

nature



SMALL WONDER

Complete skull of diminutive
dinosaur preserved in amber

Research replication

Reproducibility
lessons from DARPA's
shadow labs

Arctic ice loss

Revised picture of
mass balance for the
Greenland Ice Sheet

Extreme life

How microorganisms
survive in the lower
oceanic crust

Vol. 578, No. 7798
nature.com

End chronic kidney disease neglect

It is unacceptable that kidney-dialysis technology has changed little in the past five decades.

Dialysis almost immediately saved lives when it was invented in the first half of the twentieth century to treat kidney disease by safely accessing a patient's blood supply and filtering toxins normally removed by the kidneys.

The original dialysis machine, a rudimentary contraption invented by physician Willem Kolff in the early 1940s, was made from cellophane tubes and a wooden drum. Although today's machines are manufactured industrially, operating technologies have changed little since the 1960s. And that, as we report in a Feature on page 186, is a problem.

The World Health Organization estimates that, each year, around 1.2 million people worldwide die from kidney failure. This is partly a result of the number of people with high blood pressure and diabetes, which strain and damage the kidneys. But a combination of dialysis technology's practical limitations and affordability also means that fewer than half of the people who need treatment can access it.


In Africa, just 16% of people with kidney disease get dialysis, and even fewer can sustain the cost of treatment for more than a few months. In the United States, where dialysis can cost up to US\$91,000 per patient per year, fewer than half of those on the most common form of dialysis survive for more than five years from the onset of kidney failure.

A fundamental problem is that dialysis involves connecting patients to machines that can weigh more than 100 kilograms. For most people, that necessitates regular visits to a hospital or dialysis clinic. A typical treatment regime can take 12 hours, spread over 3 weekly sessions that see toxins filtered from the blood and levels of various salts and minerals recalibrated. For patients, this is an ordeal that is both energy-sapping and time-consuming.

Dialysis also consumes resources. Between 120 and 240 litres of filtered water are needed for each 4-hour session. By one estimate, the annual requirements of the dialysis provided around the world include more than 156 billion litres of water and roughly 1.62 billion kilowatt hours of power – roughly equivalent to the electricity needed to power a small European city for a year. Dialysis also generates some 625,000 tonnes of plastic waste.

Solutions include making dialysis more portable – so that it can be carried out at home or on the move – and finding ways to do it with less water and power. This would be of particular benefit to patients in developing countries.

Promising technologies are being developed to make the machines smaller and more portable. But it is not clear whether these will reach many of those who need them.

 Fewer than half of the people who need treatment can access it."

Despite the clear need for innovation, research funding is low. Last year, the US government and the American Society of Nephrology announced a plan to raise \$250 million over 5 years for research. This is welcome, but so far the collaboration, called Kidney X, has disbursed only \$1.1 million in grants.

The reality is that chronic kidney disease needs the kind of coordinated global effort, involving funders, researchers and patient groups, that some other conditions attract. Together, these groups must support more research into the mechanisms that underlie kidney disease, and approaches to prevent its development and progression.

At the same time, countries could encourage patients' family members to consider kidney donation. On average, transplant patients live longer than do those on dialysis – reducing the heavy health and economic burdens of this neglected killer.

DARPA 'lookalikes' must ground their dreams in reality

The US Defense Advanced Research Projects Agency knows that its freedom to invent comes with responsibility.

The government of UK Prime Minister Boris Johnson is racing ahead with plans for an Advanced Research Projects Agency (ARPA), modelled on the US original. The country is looking to boost technological competitiveness as it withdraws from the European Union. Precise details of its ARPA plans are yet to be revealed, but the available funding is expected to come to around £800 million (US\$1 billion) over 5 years.

The US Defense Advanced Research Projects Agency (DARPA), which supports ambitious technologies for military objectives, was launched in 1958 by president Dwight Eisenhower. The impetus was the Soviet Union's 1957 launch of the first artificial satellite, Sputnik, which demonstrated a level of technological prowess that shocked Western nations. Eisenhower's ambition for DARPA – established in the same year as NASA – was that the US military would never again be left behind in this way.

DARPA's best-known investments include research on the first global satellite-navigation system (known as Transit), stealth aircraft and the Internet's precursor, ARPANET. Today, the many projects funded by the agency include work on developing treatments to regrow severed limbs.

DARPA spends about \$3.5 billion a year, which is less than 1% of the total US public and private research and development budget. It's a small enough proportion to justify DARPA's reputation for taking on riskier ideas and having

a higher tolerance for failure than conventional funding agencies. One idea that famously didn't make it was the hafnium bomb, based on the flawed belief that large amounts of energy can be released by bombarding the isotope hafnium-178 with X-rays.

Although DARPA's research programme managers have relatively more flexibility in what research to fund and how, no country has been able to replicate the scale of the agency's success. Even in the United States its achievements remain unrivalled. A different advanced research projects agency for new energy technologies (ARPA-E) – launched in 2009 – is under constant threat of being eliminated by the administration of President Donald Trump.

One reason why DARPA is so hard to replicate, says DARPA historian Sharon Weinberger, is because the agency's projects have a resource that the others lack. "They have a customer with the deepest pockets in the world," she says. The US Department of Defense's annual budgets for research and procurement, totalling \$190 billion, enable it to fund successful prototypes on a large scale, to test whether they might be commercially viable.

Risk management

A closer look at DARPA shows how its managers pursue bold ideas while controlling risk. In a Comment article on page 190, members of a team working with – and in – the agency's Biological Technologies Office in Arlington, Virginia, report on an initiative launched in 2016. This assigns an independent validation team to projects to troubleshoot and reproduce research proposals. This 'shadow team' meets with the 'performing team' to learn the precise protocols and establish the necessary conditions to reproduce projects, and the two groups make joint presentations to the programme manager on progress.

The work is hard – one project took as long as 20 months to reproduce. It is also expensive: it costs between 3% and 8% of a programme's funds to make sure the technologies work. But programme managers say it is worth the investment, and the model demonstrates a more careful side to the agency than DARPA's daring image tends to evoke.

These efforts are instructive, both for dreams of a UK ARPA and for science overall. Some of UK ARPA's supporters would like to see cutting-edge technologies developed within 15 years – and a certain ruthlessness when it comes to axing the least promising ones. But an ambitious technology goal in, say, regenerative medicine or remote sensing will probably need longer before careful study can make the promises – and risks – clear.

Researchers, their managers in universities, and funding agencies all understand why effective due diligence is essential to projects. But it can be difficult for these voices to be heard when no less than the prime minister's office celebrates ARPA as "high-risk, high-pay-off research", and characterizes bureaucracy as "form-filling".

Any nation looking to replicate DARPA must realize that you can't reap the rewards of high-risk research without investing in meticulous preparation and verification. The freedom to pursue bold ideas comes with added responsibility.

“
Preparations
for coming
swarms
need to be
accelerated.”

A lack of locust preparedness

Locusts are causing a food crisis that can no longer be ignored.

While all eyes are on the coronavirus outbreak, an under-reported emergency is threatening food, health and jobs on three continents. For the past several months, swarms of the desert locust *Schistocerca gregaria* – some swarms the size of cities – have devoured crops in East Africa, the Middle East and south Asia. Some 20 million people are facing a food crisis.

Governments have been left under-prepared for the scale of these attacks, and the Rome-based United Nations Food and Agriculture Organization (FAO) has appealed for US\$138 million in urgent funding – some of which is needed to lease aircraft that can drop chemicals to curb the spread.

Locusts are an annual fixture after the rainy season – laying their eggs in moist soils. But the size of this year's swarms – the biggest for at least 25 years – are due in part to unseasonal and often torrential rains in many areas, including Eritrea, Ethiopia, Iran, Pakistan, Saudi Arabia and Yemen. The FAO's calls must be heeded – especially as more rains are to come, bringing the potential for yet more devastation. But at the same time, some of the governments concerned must ask themselves what more they could be doing to limit the damage.

In many countries, meteorological offices share climate and weather data with what are called desert locust-control offices. These are set up to forecast locust infestations, and to advise on potential crop losses and mitigation measures. East Africa has a regional body called the Desert Locust Control Organization for East Africa, headquartered in Addis Ababa and funded by nine African countries.

Nature has been told that some members – such as Djibouti, Somalia and Sudan – have been unable to pay their membership fees for many years and collectively owe the organization more than \$8 million. Uganda, which partially cleared its arrears last month, still owes \$2 million. Somalia and Sudan have both experienced severe conflict, so it's understandable that locusts have not been a priority. But the insects can be just as threatening to well-being, and if individual countries can't pay their way, then the African Union or the UN need to step in. Paying into locust-control offices should be regarded as keeping up an insurance policy. The hope with insurance is that it's never needed, but the facility must always be there should the need arise.

The focus now is rightly on emergency food relief. But preparations for coming swarms need to be accelerated. The African Union and the UN must ensure that countries' desert locust organizations, informed by the latest research, are better equipped to help when the time comes.

World view

Want to do better science? Admit you're not objective



By Angela Saini

When science is viewed in isolation from the past and politics, it's easier for those with bad intentions to revive dangerous and discredited ideas.

One of the world's leading universities – University College London (UCL) – has completed an inquiry into its support for the discredited pseudoscience of eugenics. Funds linked to Francis Galton, a racist who believed it was possible to improve the British population through selective breeding, and who founded the Eugenics Records Office at UCL in 1904, continue to line the university's coffers to the value of more than £800,000 (US\$1 million).

The inquiry's report, released on 28 February, recommended renaming lecture theatres and buildings bearing Galton's name and that of another prominent geneticist. Although this is welcome, it does not acknowledge just how much yesterday's mistakes survive in modern science.

As I found while writing my 2019 book *Superior: The Return of Race Science*, geneticists today rightly treat eugenics as a laughable proposition, and the concept of biological race – the belief that humans can be subdivided into distinct groups with meaningful differences between them – as easily debunked nonsense. But this ignores how these ideas manifest in the real world. They can only be truly understood as age-old intellectual threads, embedded in politics as much now as ever.

In failing to recognize that science can be political, the scientific community allows the resurrection of dangerous ideas. Acting as if theories – especially those about humans – exist in cultural or political vacuums is a ridiculous fallacy.

The UCL inquiry was prompted in part by 2018 revelations that a now-former honorary fellow had been booking meeting space for secretive conferences discussing race and eugenics. Many people – even members of the inquiry committee – are concerned that the investigation did not go far enough in connecting the pseudoscience of the past with ongoing attempts to keep that pseudoscience alive.

In the same month that UCL released its report, news broke that Dominic Cummings, a self-proclaimed science enthusiast and special adviser to the UK prime minister, had hired an aide who espouses eugenicist views. Now resigned, Andrew Sabisky had suggested compulsory contraception to halt the growth of a “permanent underclass”.

When a survey conducted as part of the UCL inquiry asked staff and students whether “we should separate science and politics”, it found agreement among higher percentages of those in the sciences and engineering than

Racism and prejudice are woven into the structures in which we all live and work – and into us.”

in the social sciences and history. In my coverage of the inquiry, I've seen that it was not the university's biologists, but its humanities scholars – including curator Subhadra Das and historian Joe Cain – who forced their workplace to confront a sordid history that some geneticists had been willing to overlook.

“If the past is to be called upon to legitimize the present, as it so frequently is, then the veracity of such a past has to be continuously vetted,” writes Romila Thapar, a historian at Jawaharlal Nehru University in New Delhi. Those who seek to bring back eugenics prefer to gloss over the past and treat the concept as a purely scientific proposition. This is made easier by scientists who are willing to separate science from politics, even when – as with eugenics – it was inherently political to begin with.

The UCL survey also revealed that members of the university community who are disabled or from minority ethnic backgrounds are more likely to feel that the legacy of eugenics is still present than are those who are white and not disabled. The people on the receiving end of the world-view that drove eugenics understand how alive it remains.

Scientists who imagine that bias lies in others, not themselves, fail to recognize that to live in the world today is to be drip-fed assumptions and prejudices that guide our thoughts and actions. If it were any other way, the demographics of academia would be more equitable, and the current strain of genetic determinism in governments wouldn't be possible. Racism and prejudice are woven into the structures in which we all live and work – and into us.

A lack of education means we fail to draw these links. At a philosophy festival last September, I spoke about non-European cultures and their contributions to science and mathematics. One scientist remarked that he had no need to know about what had been done in ‘bongo bongo’ land. The audience was shocked; I was disappointed. He was a product of a system that had not taught him that he needed to know better. It is this system we need to fix.

Scientists rarely interrogate the histories even of their own disciplines. When I studied engineering at university, I was expected to write just one essay on ethics in four years. No wonder that new technologies perpetuate racial and gender stereotypes, or that automated facial recognition struggles to identify people with darker skin.

The best research is done not when we pretend that we are perfectly objective, but when we acknowledge that we are not. The UCL inquiry report recommends that students and staff be exposed to the history of eugenics, and that students be encouraged to value the history of their own fields. I would go further. Scientists need both history and the social sciences to develop the intellectual tools to think critically about their research and how it affects society. This isn't just helpful – it's vital.

Angela Saini is a science journalist and broadcaster based in London. www.angelasaini.co.uk

News in brief

CLIMATE CHANGE CONTRIBUTED TO AUSTRALIA'S EXTREME BUSH-FIRE WEATHER

Global warming increased by at least 30% the risk of the weather conditions that drove the recent catastrophic bush fires in Australia, says a group of climate scientists who analysed the disaster.

Australia experiences bush fires regularly, but the latest event was unprecedented in its severity and scale.

The World Weather Attribution (WWA) project sought to measure climate change's contribution to the bush-fire conditions seen in southeast Australia. Fires in the region were particularly severe, and killed dozens of people and destroyed thousands of homes.

The group assessed bush-fire conditions with an index that tracks 'fire weather'. This calculates the chance of fire in a location on the basis of variables such as temperature, humidity, wind and rainfall. The analysis did not consider non-weather factors, such as how a fire started.

"This is a highly conservative assessment," says David Karoly, a climate scientist based in Melbourne at the Commonwealth Scientific and Industrial Research Organisation, who was not

involved in the analysis.

The project's researchers also say that the result is conservative. Models have mostly underestimated the rise in temperatures that has been observed since the Industrial Revolution, says Geert Jan van Oldenborgh, a climate researcher at the Royal Netherlands Meteorological Institute in De Bilt, the Netherlands, and co-author of the WWA analysis.

The team also examined whether climate change had influenced two of the components that are used to measure fire weather: extreme temperatures and drought. The results, which were posted on the group's website and have not been peer reviewed, suggest that human activity doubled the chance of heatwave conditions during the fires, but do not show that climate change contributed to the extremely dry conditions that Australia experienced.

Climate change definitely played a part in the fires, says Andy Pittman, a climate scientist at the University of New South Wales in Sydney, Australia. But he questions whether the results are meaningful, because models struggle to simulate fires.



EU CHAMPIONS BOLD NEW CLIMATE LAW

The European Commission has proposed a law that would give it far-reaching power to dictate the course of political action on climate change in the European Union.

The draft climate law, unveiled in Brussels on 4 March, would create a legally binding commitment for the EU to reduce its greenhouse-gas emissions to net zero by 2050 – which means any emissions would have to be offset by greenhouse-gas uptake, for example by trees or through carbon capture and storage technologies.

The proposal would also give the commission power to set binding short-term climate targets that don't need unanimous approval from all 27 member states. Policy analysts say that some countries could strongly oppose these measures, so the current draft is unlikely to be approved without substantial amendments.

The draft does not mention specific EU emissions-reduction milestones between now and 2050. Climate campaigners – including Swedish teenager Greta Thunberg (pictured, right, with commission president Ursula von der Leyen) – say the law doesn't go far enough. "We don't just need goals for 2030 or 2050. We, above all, need them for 2020 and every following month and year to come," a group of 34 young activists wrote in an open letter to EU leaders.

CORONAVIRUS: CHILDREN AS SUSCEPTIBLE AS ADULTS

Children are just as likely as adults to become infected with the new coronavirus, according to a detailed study on the spread of the virus. The analysis, based on data from Shenzhen in China, provides a partial answer to a key question surrounding the outbreak: the role of children.

Previous studies have suggested that children are less likely than other age groups to develop severe symptoms when infected by the coronavirus. But it was not clear whether this was because they weren't getting infected or because they were fighting off the infection more effectively.

"Kids are just as likely to get infected," says Justin Lessler, an infectious-disease epidemiologist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, who co-led the study, which was posted to the medRxiv preprint server on 4 March (Q. Bi *et al.* <http://doi.org/dpf9>; 2020).

The researchers followed 391 people diagnosed with coronavirus, and 1,286 of their close contacts to see whether these contacts tested positive for the virus even if they didn't show symptoms. Overall, the team found that children under 10 who had potentially been exposed to the virus were just as likely to become infected as other age groups, with between 7% and 8% of contacts of known cases later testing positive.

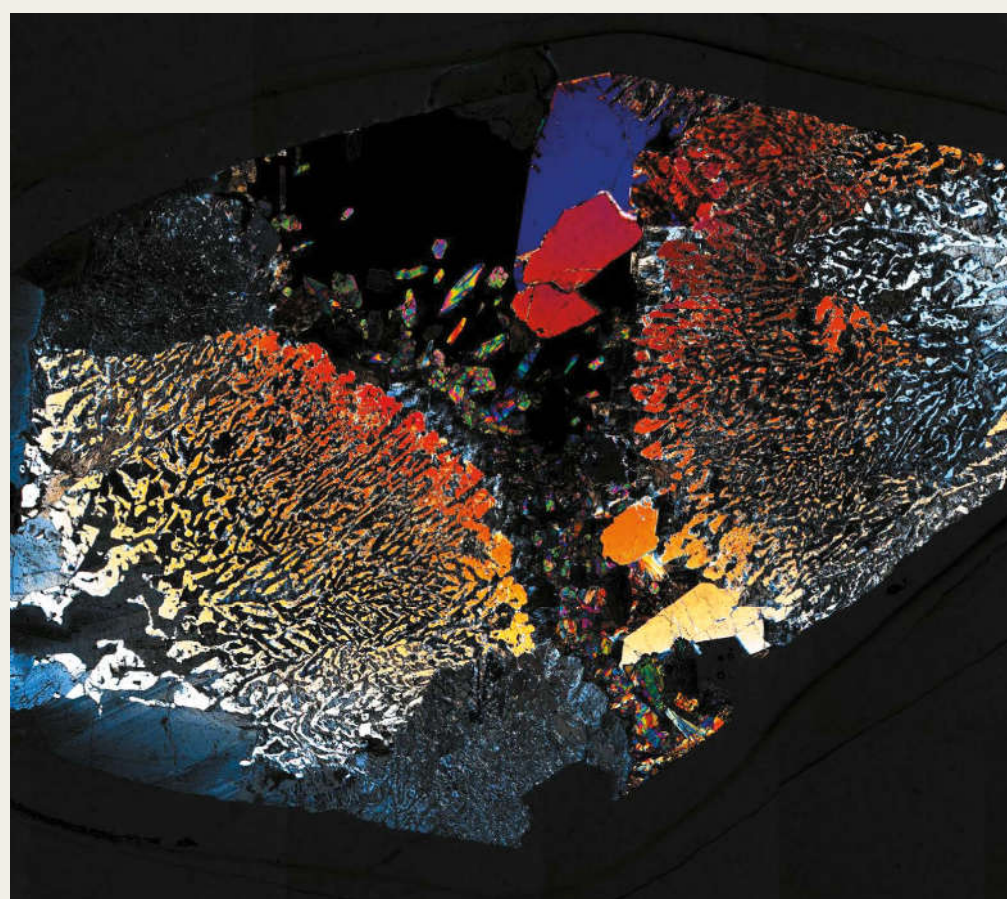
The findings could influence measures intended to halt the spread of the virus. "This is a key piece of data that may support school closures as an effective intervention," Caitlin Rivers, an epidemiologist at Johns Hopkins Bloomberg School of Public Health, said in a tweet on 5 March.

TINY DINOSAUR PRESERVED IN AMBER FOR 100 MILLION YEARS

A creature exquisitely preserved in amber for 100 million years is the smallest known dinosaur of its era.

The animal's skull, described this week in *Nature*, is less than 2 centimetre long – and suggests the dinosaur was about the size of a bee hummingbird (*Mellisuga helenae*), the smallest living bird (L. Xing *et al. Nature* 579, 245–249; 2020). Researchers put the new dinosaur in a genus they called *Oculudentavis*, meaning 'eye-teeth bird'. "It reveals to us a whole new lineage of birds," says Jingmai O'Connor, a palaeontologist at the Chinese Academy of Sciences Institute of Vertebrate Paleontology and Paleoanthropology in Beijing, who co-led the study.

The fossil, which was found in Myanmar and comes from the Mesozoic era, is exceptionally well preserved for a specimen of its size. Its beak is crammed with dozens of sharp teeth, suggesting that in life, the creature preyed on insects and other small invertebrates. Its eyes protrude from either side of its skull so, unlike most modern predators, this dinosaur did not have binocular vision. And its size and age mean that miniaturization in birds occurred earlier than scientists previously thought.



Microbes found deep in Earth's crust

This thin slice of rock, viewed using a microscope, is part of a sample taken from the Atlantis Bank at the bottom of the Indian Ocean, where researchers have found microbes living deep within Earth's oceanic crust.

A team led by marine microbiologist Virginia Edgcomb at Woods Hole Oceanographic Institution in Massachusetts identified a range of bacteria, fungi and archaea that inhabit hairline cracks in rocks up to 750 metres below the ocean floor, where they are able to live and grow despite extremely limited resources (J. Li *et al. Nature* 579, 250–255; 2020). By measuring enzyme activity and gene expression, the researchers showed that these microbial communities have adapted to their conditions by maintaining low levels of cellular activity and feeding on carbon from fragments of amino acids and other organic molecules carried by deep ocean currents.

News in focus



DAVID RYDER/REUTERS

Medics tend to a patient at the Life Care Center of Kirkland, a facility linked to several of Washington state's confirmed coronavirus cases.

THE RACE TO UNRAVEL THE BIGGEST CORONAVIRUS OUTBREAK IN THE UNITED STATES

As cases in Washington state soar, virologists are working around the clock to diagnose cases, reveal routes of transmission and test treatments.

By Amy Maxmen
in Seattle, Washington

Rohit Shankar left the virology laboratory at 2 a.m. one night last week, and was back at the lab bench by 7 a.m. the same day. "It's okay," he says, "I had a doughnut and a coffee."

Shankar, a medical scientist, and his colleagues at the University of Washington in Seattle are poised to trigger an exponential rise in the number of cases of the coronavirus disease COVID-19 confirmed in and around the city, in western Washington state. That's because

last week, they began analysing a mountain of nose and throat swabs collected from hospitals. Already, the researchers are seeing clear signs that the virus has infected vastly more people than have been formally detected.

Washington state has become the United States' ground zero for COVID-19, which has now spread to more than 100 countries. Washington has declared a state emergency, and ten people there have died from the disease. But the number of confirmed cases in Washington — around 160 as *Nature* went to press — is an underestimate resulting from a lack of testing, researchers agree. A genomic analysis posted

online on 29 February suggested that hundreds of people in western Washington might be infected (see go.nature.com/3clwou3). Academic scientists have mostly been prevented from measuring the extent of the US outbreak because of federal restrictions on the number of labs qualified to run diagnostic tests. But that is changing.

Dozens of virologists and genomicists have now kicked into high gear in Seattle, dropping or adapting projects to devote resources to the outbreak. Researchers are working around the clock to find out how many people have the disease in the area. Others are analysing

News in focus

genomes to reveal how the virus is transmitted, or developing new therapies. The scientists are racing to help Washington avoid the fate of Hubei province in China, where more than 3,000 people have died of COVID-19 so far. The coronavirus emerged in the province's city of Wuhan in December, and the initial response from officials was slow.

"We are past the point of containment," says Helen Chu, an infectious-disease specialist at the University of Washington School of Medicine (UW Medicine) in Seattle. "So now we need to keep the people who are vulnerable from getting sick."

Working in shifts

"I remembered to eat around 10 p.m. last night," says Keith Jerome, director of the clinical virology department at UW Medicine, nodding to an empty pizza box on a conference-room table. In January, his group quickly adapted a PCR test described by the World Health Organization that identifies snippets of the virus's genome sequence.

But when COVID-19 reached the United States, his team couldn't check the accuracy of its test – done by analysing samples from people known to have the disease – because of regulations set by the US Food and Drug Administration (FDA) to ensure that tests are accurate before clinics rely on them. But researchers became impatient with the pace of FDA processes as the outbreak hit home. On 29 February, the agency announced that it would allow certain academic labs to test people for the disease – opening Jerome's lab to a flood of samples.

The lab is now helping hospitals and the state to clinically diagnose cases of COVID-19. Jerome estimates that with its high-throughput molecular-biology equipment, the group can screen around 1,000 samples per day – roughly 5 times as many as the state department of health's lab. "That's why it's such a problem to not have labs like this involved in the early days of an outbreak," Jerome says. Now that they're rolling, the researchers have plans to bulk up their capacity to 4,000 samples daily. Jerome says they're quadrupling their equipment thanks to university funds, asking other labs in the city to lend them PCR machines, and adding more researchers to their ranks. "We're making plans for different shifts around the clock so that we don't burn people out," he says.

Because the test is still new, the lab sends positive samples to the state lab for confirmation. Then the state alerts the US Centers for Disease Control and Prevention (CDC), which is responsible for official US case counts.

Undetected spread

Meanwhile, in Chu's lab, boxes containing swab kits balance precariously on lab benches and carts. Her group has just received the green

light from the FDA and the CDC to analyse nose swabs for the coronavirus. It's for a scientific study, rather than as part of the state's public-health response. Still, the researchers act on the results in real time. If samples test positive, they notify the health department.

The project stems from the multi-institution Seattle Flu Study, which Chu has co-led since 2018. In that study, participants who feel as if they have a cold or influenza swab their nose and send the sample to the lab, where

"Our emergency rooms are being flooded today by people saying they want a test."

researchers sequence any influenza-virus genomes they contain. Analyses of these genomes reveal the trail that the flu takes as it passes around households, homeless shelters, offices and communities in the city.

As COVID-19 overwhelmed China in January, the Seattle Flu Study researchers felt sure that the epidemic would soon spread to the United States – and that they should track that, too. So they customized a genetic test, similar in design to what they had used for flu viruses, so that it could identify genetic sequences of the new coronavirus. On 27 February, soon after they had received clearance to use it, their 'COVID panel' detected the virus in a swab from a teenager north of Seattle. His case would prove pivotal because he hadn't travelled internationally. The team alerted the county health department, and set about sequencing the whole genome of the virus.

After barely 24 hours, the researchers

had sequenced the genome. They posted the sequence to an online platform called GISAID. Then a collaborator at Nextstrain, an online project that visualizes the spread of viruses through genomic analyses, compared the genome with dozens of others that had been sequenced around the world. In a series of tweets on 29 February, Nextstrain's co-founder, computational biologist Trevor Bedford, explained his findings. The sequence contained an unusual genetic variation that matched that of the virus from the first person reported to have COVID-19 in the United States, on 20 January – a man treated at a hospital north of Seattle. This meant that the virus had probably been circulating around western Washington for six weeks. Bedford calculated that in that time frame, up to 1,500 people could have been infected.

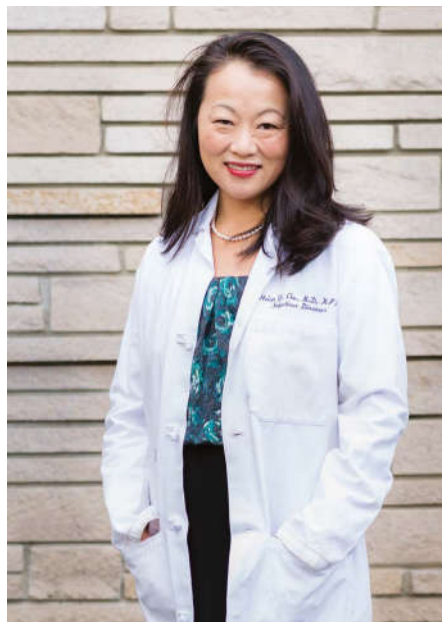
In the days that followed, officials in Seattle reported a rising number of deaths from COVID-19 among older people who hadn't travelled. Yet what is lacking, says Cassie Sauer, chief executive of Washington State Hospital Association, are diagnostic tests. "Our emergency rooms are being flooded today by people saying they want a test," she says. Even with the added power at the University of Washington, supplies are limited.

'Spending money like it's nothing'

Some of this burden could soon be relieved as the Seattle Flu Study changes course. People who feel ill but aren't in dire need of hospital care can now order a swab kit from the study website. "We're ramping up to screening 570 tests a day, and hope to do 1,000 per day in a few weeks," says Lea Starita, a genomicist on the project at UW Medicine. She's scrambling to find researchers to join the lab and is purchasing new tools. "I've been spending money this week like it's nothing," she says, listing US\$390,000 in equipment and reagents off the top of her head.

As the swab kits pile up in Chu's lab, she mentions more than 2,500 samples collected this year for the flu study. Some might contain the new coronavirus, and genomic analyses could reveal how it circulated undetected around Seattle. But with the ever-mounting workload, there's no time to analyse them now. Chu is also involved in an effort to isolate antibody proteins from people with COVID-19, in the hope that researchers can develop a treatment. And she's leading a clinical trial at UW Medicine to see whether the experimental antiviral drug remdesivir could be used to treat the disease.

Chu doesn't spare a thought for the work she's put on hold. She's in triage mode, prioritizing the most urgent questions. "As diagnostic testing ramps up, it will become clear that this is everywhere," she says. Seattle's scientists might help Washington to mitigate the outbreak's harm to lives and the economy, and provide models for other states and countries.



Helen Chu is studying coronavirus spread.

JULIE NIMMERGUT

MONKEYS AND MICE ENLISTED TO FIGHT CORONAVIRUS

Animal models can reveal how infections develop, and aid drug and vaccine efforts.

By Ewen Callaway

With no sign that the coronavirus is going away, researchers are looking to animals to understand COVID-19. They are testing monkeys, mice and even ferrets to answer key questions about the disease and to fast-track potential drugs and vaccines for clinical trials.

Teams in China have reported initial findings from studies in which they infected monkeys and mice engineered to be susceptible to infection by the coronavirus, called SARS-CoV-2. And a team at the Australian Animal Health Laboratory in Geelong is studying the infection in ferrets, before testing potential vaccines. Ferrets are a popular model for respiratory infections because their lung physiology is similar to humans'.

But no animal model is perfect. "There's going to be a need not just for one animal model, but multiple," says David O'Connor, a virologist at the University of Wisconsin-Madison.

Mild illness

O'Connor and fellow University of Wisconsin virologist Thomas Friedrich are part of a network of about 60 scientists who are sharing details of their efforts to study the infection in primates and other animals. They were excited to read about experiments in non-human primates infected with COVID-19, reported in a preprint on 27 February (C. Shan *et al.* Preprint at Research Square <https://doi.org/10.21203/rs.2.25200; 2020>).

That research, led by virologist Chao Shan at the Chinese Academy of Sciences Wuhan Institute of Virology, found that rhesus macaques infected with the coronavirus had a fairly mild illness. None developed fevers, but X-rays of their lungs showed signs of pneumonia similar to those in humans with COVID-19. The researchers also monitored two animals for three weeks; these monkeys lost weight, but didn't have other serious symptoms.

The fact that the monkeys seemed to develop symptoms similar to those in people with mild forms of COVID-19 is an important takeaway, says O'Connor. But to find models for severe human infections, researchers will



Rhesus macaques are used to study the virus.

have to look at different animals and vary experimental factors such as the route by which the virus is administered, he adds.

Many researchers are turning to lab mice to test drugs and vaccines, and to investigate the nature of the infection. But ordinary mice seem to be resistant to infection by SARS-CoV-2. So researchers are hoping to use mice bred to produce a human version of the protein ACE2, which the virus uses to enter cells.

One lab has already begun infecting them with coronavirus. A team of researchers in China found that, like rhesus monkeys, the mice seemed to develop only mild illness, showing weight loss and signs of pneumonia but nothing more severe (L. Bao *et al.* Preprint at bioRxiv <http://doi.org/dph2; 2020>). Qin Chuan, a virologist at Peking Union Medical College in Beijing who co-led the study, says that, in unpublished work, his team also identified several drugs that slowed the virus's replication and limited the animals' weight loss.

Animals that develop mild infections could be useful for testing drugs and vaccines, says Stanley Perlman, a coronavirologist at the University of Iowa in Iowa City whose lab developed an ACE2 mouse strain. But he's thinking about developing other mouse models to better mimic severe cases. "Models are imperfect; we do the best we can," Perlman says.

Why does the coronavirus spread so easily?

A microscopic feature could make the virus more infectious than the SARS virus.

As the number of coronavirus infections passes 100,000 worldwide, researchers are racing to understand what makes it spread so easily.

A handful of genetic and structural analyses have identified a key feature of the virus — a protein on its surface — that might explain why it infects human cells much more readily than does the coronavirus that causes severe acute respiratory syndrome, or SARS.

Other groups are investigating the doorway through which the new coronavirus enters human tissues — a receptor on cell membranes. The cell receptor and the virus protein offer potential targets for drugs to block the pathogen, but researchers say it is too early to be sure.

To infect a cell, coronaviruses use a 'spike' protein that binds to the cell membrane. Genomic analyses of the new coronavirus have revealed that its spike protein differs from those of close relatives, and suggest that the protein has a site that is activated by an enzyme called furin.

This is significant because furin is found in many human tissues, including the lungs, liver and small intestine, which means that the virus has the potential to attack multiple organs, says Li Hua, a structural biologist at Huazhong University of Science and Technology in Wuhan, China, the city where the outbreak began.

The finding could explain some of the symptoms observed in people with the coronavirus, such as liver failure, says Li, who co-authored a genetic analysis of the virus (H. Li *et al.* Preprint at <http://chinaxiv.org/abs/202002.00062; 2020>).

Other groups have also identified the activation site as possibly enabling the virus to spread easily between humans.

But some researchers are cautious about overstating the role of the site. "We don't know if this is going to be a big deal or not," says Jason McLellan, a structural biologist at the University of Texas at Austin, who co-authored another structural analysis (D. Wrapp *et al.* *Science* <http://doi.org/ggmtk2; 2020>).

By Smriti Mallapaty



Litigant Nathan Baring speaks during a meeting of the US Senate Climate Change Task Force.

CLIMATE LAWSUITS BREAK NEW GROUND TO PROTECT THE PLANET

Despite recent defeats, activists are optimistic that courts will provide relief from climate change.

By Giuliana Viglione

Twenty-year-old Nathan Baring is a third-generation Alaskan. Within his lifetime, he has seen winters shorten, cod fisheries collapse and cultural traditions suffer. He grieves for an Arctic that is disappearing before his eyes. “There is a very distinct loss of place here,” he says.

Baring decided to take action. He is one of 21 young plaintiffs who, in 2015, filed a lawsuit seeking to force the US government to reduce greenhouse-gas emissions. A federal appeals court dismissed the case, known as *Juliana v. United States*, in January.

But other attempts to fight climate change in the courts have been more successful. Two weeks ago, the UK Court of Appeal blocked plans to build a third runway at Heathrow Airport, saying that allowing the expansion would violate the country’s obligations to the Paris climate agreement.

Such decisions are inspiring and instructing activists and municipalities around the world. But as litigants fight scores of such cases, one thing is clear. “There is no silver litigation bullet for climate change,” says Michael Gerrard, director of the Sabin Center for

Climate Change Law at Columbia University in New York City.

On 18 February, the International Bar Association released a model for how to litigate climate change, laying out legal arguments and precedents that might help future plaintiffs.

What is likely to succeed seems to vary around the world. And as plaintiffs learn from their experiences in the courts, they are adjusting their tactics. The Heathrow case is the first

“There is no silver litigation bullet for climate change.”

major ruling based on the Paris agreement and could spur more suits that rely on those obligations. In other parts of the world, plaintiffs are increasingly focusing on seeking damages from polluters themselves.

The *Juliana* case is one of more than two dozen brought around the world on the basis of ‘public trust’ arguments, which say that the state has a duty to protect public resources from harm. Such arguments are closely linked to the idea that the fundamental right to life

is inextricably tied to a healthy environment.

Although the US court found that the plaintiffs had been harmed by the government’s inaction on climate change, the judges ultimately ruled that it was not within the court’s power to legislate climate policy.

But outside the United States, the human-rights approach is the legal strategy that has had the most success in forcing governments’ hands, says John Knox, an expert on international environmental law at Wake Forest University in Winston-Salem, North Carolina.

In *Urgenda Foundation v. State of the Netherlands*, a case brought in 2015 by an environmental group and nearly 900 Dutch citizens, the Dutch supreme court mandated that the government achieve a 25% reduction in greenhouse-gas emissions from 1990 levels by the end of 2020 to protect its citizens from the harms of a warming climate system.

And in *Demanda Generaciones Futuras v. Minambiente*, Colombia’s Supreme Court ordered the government to implement protective measures to halt deforestation in the Amazon – that case was brought by 25 young Colombians. Like *Juliana*, both cases rested on the idea that the right to life is endangered by threats to the environment.

Their success suggests that we can expect to see more suits brought by citizens against their governments, says Knox. Meanwhile, climate activists are watching closely to see how these governments comply with the court-ordered actions.

Damage limitation

In the United States, the scope of *Juliana* – which sought to force the US government to not only stop permitting and subsidizing fossil-fuel use, but also implement a plan for reducing atmospheric carbon dioxide levels – meant it was “always going to be a long shot”, says Ann Carlson, who studies environmental law at the University of California, Los Angeles. And given the increasingly conservative make-up of many US federal courts, Carlson thinks it is unlikely that future cases based on similar arguments will succeed.

Instead, Gerrard and Carlson expect more US activists and municipalities to pivot from targeting the government to suing the producers of emissions themselves. This strategy is seen as more pragmatic because cash penalties can, in some cases, be used for climate mitigation.

In *County of San Mateo v. Chevron Corp.*, several Californian cities and counties are seeking funds from major fossil-fuel corporations to fund infrastructure for adaptation to sea-level rise. Oral arguments in the latest appeal were heard on 5 February, but a ruling has yet to be handed down. Several other individuals and localities, including the state of Massachusetts, are currently suing ExxonMobil and other companies for allegedly deceiving

MARK WILSON/GETTY

consumers about the risks of fossil-fuel use.

Similar arguments for compensation are also being made outside the United States. In November 2015, a Peruvian farmer named Saúl Lliuya brought a case in German courts against the German utilities company RWE, the largest emitter of CO₂ in the European Union.

Lliuya, who lives near a glacial lake, alleges that RWE's emissions are partially responsible for the dangerously high water levels seen at the lake as nearby glaciers have melted. He is seeking 0.47% of the costs of flood-protection measures for his town, equal to RWE's proportion of global CO₂ emissions from 1751 to 2010.

The case was initially dismissed, but an

appeals court has since ruled that Lliuya's complaint was admissible, and the court has ordered the parties to submit expert evidence – the first time such a case has moved to the evidentiary stage. Gerrard says a victory could spur similar lawsuits around the world.

Although the recent ruling in *Juliana* was disappointing, the plaintiffs say that they are heartened by the court's finding that they had been harmed by the government's inaction on climate change. They are now preparing to appeal, and are optimistic that they will get a chance to argue their case in front of a jury. "We have many paths forward," Baring says. "This is certainly not the end of the road for us."

the technique to edit the genomes of cells that have been removed from the body. The material is then infused back into the patient.

It's a significant jump from treating cells in a dish, says Fyodor Urnov, who studies genome editing at the University of California, Berkeley. "It is akin to space flight versus a regular plane trip," he says. "The technical challenges, and inherent safety concerns, are much greater."

"It's an exciting time," adds Mark Pennesi, a specialist in inherited retinal diseases at Oregon Health & Science University in Portland. Pennesi is collaborating with the pharmaceutical companies Editas Medicine of Cambridge, Massachusetts, and Allergan of Dublin to conduct the trial, which has been named BRILLIANCE.

Mutation eradication

This is not the first time gene editing has been tried in the body: an older gene-editing system, called zinc-finger nucleases, has already been administered directly into people participating in clinical trials. Sangamo Therapeutics of Brisbane, California, has tested a zinc-finger-based treatment for a metabolic condition called Hunter's syndrome. The technique inserts a healthy copy of the affected gene into a specific location in the genome of liver cells. Although it seems to be safe, early results suggest it might do little to ease the symptoms of Hunter's syndrome.

But the BRILLIANCE trial is the first to deploy the popular CRISPR–Cas9 technique – which has been hailed for its versatility and ease of design – directly in the body. In BRILLIANCE, gene editing is used to delete a mutation in the gene *CEP290* that is responsible for LCA10.

The condition is a particularly attractive target for a gene-editing approach. Conventional gene therapies use a virus to insert a healthy copy of the mutated gene into affected cells. But *CEP290* is too large to slip the entire gene into a viral genome, says Artur Cideciyan, who studies retinal diseases at the University of Pennsylvania in Philadelphia.

And although mutations in *CEP290* disable light-sensing cells called photoreceptors in the retina, the cells are still present and alive in people with LCA10. "The hope is that you can reactivate those cells," says Pennesi. "This is one of the few diseases where we think you could actually get an improvement in vision."

Early results from another therapy suggest that this might be the case. Cideciyan has teamed up with ProQR of Leiden, the Netherlands, to treat people with LCA10 using an experimental treatment called sepfarsen. Early results suggest that sepfarsen, which uses a technique called antisense therapy to correct an LCA10-causing mutation in RNA made from the *CEP290* gene, can improve vision in people with LCA10.

CRISPR TREATMENT INSERTED DIRECTLY INTO BODY FOR THE FIRST TIME

Experiment tests a gene-editing therapy for a hereditary blindness disorder.

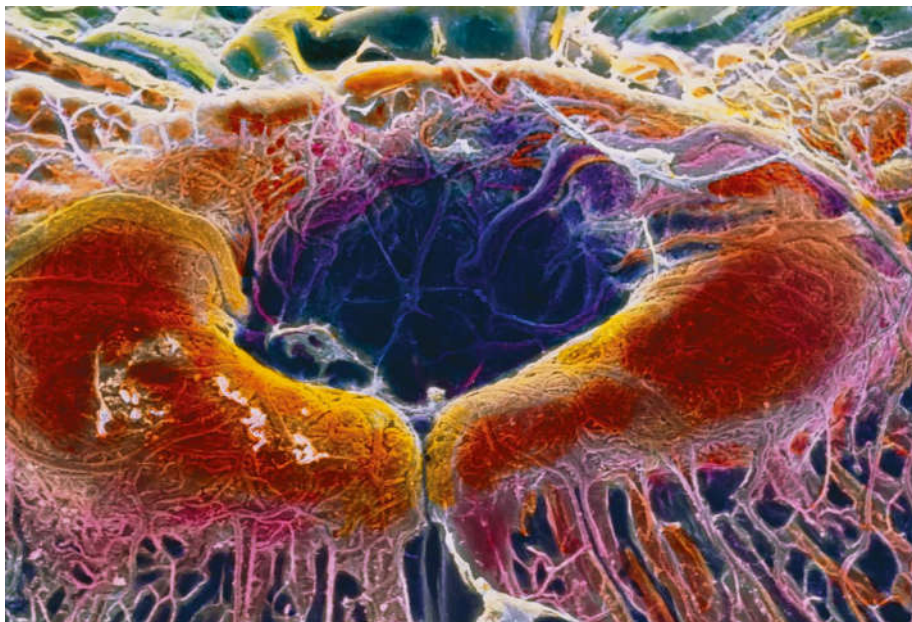
By Heidi Ledford

A person with a genetic condition that causes blindness has become the first to receive a CRISPR–Cas9 gene therapy administered directly into their body.

The treatment is part of a landmark clinical trial to test the ability of CRISPR–Cas9 gene-editing techniques to remove

mutations that cause a rare condition called Leber's congenital amaurosis 10 (LCA10). No treatment is currently available for the disease, which is a leading cause of blindness in childhood.

For the latest trial, the components of the gene-editing system – encoded in the genome of a virus – are injected directly into the eye, near photoreceptor cells. By contrast, previous CRISPR–Cas9 clinical trials have used



The human retina. A CRISPR therapy has been inserted directly into a person's eye.

TURBOCHARGING

From the start, Mat Risher swore that dialysis wouldn't upend his life. He had been working at a software company, conducting research on a car-racing simulator, when kidney damage from lupus forced him to start the blood-filtering treatments three times a week.

Five years have slipped by, and the sessions have sapped his resolve. The 33-year-old now works part-time. On good days, he enjoys trying out new recipes. On bad days, his lupus flares up and the strain of incessant dialysis leaves him drained. "The times in between [dialysis days], I have no social life, no dating life," says Risher, who lives just outside Seattle, Washington. "I have become a recluse in my room."

Risher is relatively fortunate; he has access to treatment, whereas up to seven million people could die each year without getting such care¹. But Risher, a member of a patient advisory board at Seattle's Center for Dialysis Innovation (CDI), is impatient for a more liveable option than dialysis – which has remained largely the same for 50 years.

Walk into any facility, says CDI co-director Buddy Ratner, and you'll find a big machine at the bedside of every person undergoing dialysis. "These days it's going to have LCD screens and modern controls," he says. "But look at the pictures in the 1960s of those machines. They look rather similar to what we're doing today." Survival has increased, but still, just 42% of US patients receiving the most common form of treatment, known as haemodialysis, live even for five years – shorter than for many cancers.

Ratner is among an international cadre of physicians, bioengineers and entrepreneurs who are working to revolutionize treatment for kidney failure, designing devices that are portable enough to carry into work or strap around the waist. Some are even developing artificial kidneys that could be surgically implanted.

The complexities remain daunting. Dialysis poorly mimics the sophistication of the human kidney, and improved and more portable versions will need miniaturized components and a substantial reduction in the amount of water required. Any approaches that make use of biological materials will face steep regulatory hurdles, too.

But a new wave of funding is helping to reverse the years of stagnation. Last year, US President Donald Trump issued an executive

order on kidney health, including strategies to reduce the shortage of kidneys available for transplantation, encourage more dialysis at home and incentivize research into artificial kidneys through a partnership called KidneyX. The partnership is led by the US government and the American Society of Nephrology and plans to raise US\$250 million over the next five years. Last year, it awarded a total of \$1.1 million to 15 US-based research teams tackling various pieces of the dialysis puzzle, including groups pursuing wearable dialysis devices and bioengineered kidney grafts.

Around the world, clinical trials of portable devices are advancing, and researchers are finalizing a low-tech approach that they hope will reach regions of the world where clean water is unreliable and dialysis is scarce.

All these efforts are a drop in the ocean compared with the hefty bill to treat people living with end-stage kidney disease – at least US\$35 billion annually in the United States alone. But the field is bullish. John Sedor, a nephrologist at the Cleveland Clinic in Ohio, who chairs the KidneyX steering committee, predicts that a much more portable device will be available in the next five years, and the first wearable device in the next decade. "I think this is a remarkable time and we're at a tipping point in our field," he says.

That innovation is long overdue, says Valerie Luyckx, a nephrologist at Graubünden Cantonal Hospital in Switzerland who researches the global burden of kidney disease. Dialysis "is a multibillion-dollar industry, with multiple billions in profits since the early 1960s", she says. "And nobody has bothered to try to innovate until all of the sudden there is research and grant funding for it."

A smart organ

The kidneys are complex and resilient organs, each roughly the size of a fist. They filter some 140 litres of blood each day, leaving behind a litre or two of water and waste in the form of urine.

Each kidney features a latticework of roughly one million tiny filtering units, called nephrons. Blood entering a nephron passes through a cluster of tiny vessels called the glomerulus. The thin walls of the glomerulus enable waste, water and other small molecules to pass through, while blocking larger ones such as proteins and blood cells. From there, the filtered fluid



flows into kidney tubules, where the balance of minerals, water, salts and glucose is calibrated and molecules necessary for bodily functions are reabsorbed into the bloodstream.

But many medical conditions can strain the kidneys, including diabetes, obesity and high blood pressure. And those conditions are becoming more common. By 2030, it's projected that 5.4 million people worldwide will be getting dialysis or a transplant, and many more will die without¹.

For haemodialysis, patients usually need to travel to a clinic, where they are connected to a machine weighing more than 100 kilograms that filters the patient's blood through a semi-permeable membrane, designed to replicate the function of the glomerulus. Then a water-based dialysis solution is used to

MOHAMMED HUWAIS/AP/GETTY

IG DIALYSIS

After decades of slow progress, researchers are starting to test better treatments for people with failing kidneys.

By Charlotte Huff



Kidney dialysis is inefficient and exhausting for patients such as this person in Yemen.

having dropped out of school.”

Still, there has been a lack of drive to improve the procedure, in part because the treatment has proved highly profitable for dialysis providers around the world, says Murray Sheldon, a physician and associate director for technology and innovation at the Center for Devices and Radiological Health at the US Food and Drug Administration (FDA). “They have a cash cow. And there’s no need to do any innovation.” Dialysis companies challenge that argument. Brad Puffer, a US-based spokesperson for Fresenius Medical Care in Waltham, Massachusetts, says that his company is investing in improvements, including a haemodialysis device that incorporates a material designed to reduce blood clotting, a potential side effect that today’s recipients have to take medication to prevent.

Kidney in a backpack

One of the big problems with modern dialysis is that the machines require vast amounts of water: 120–180 litres for each 4-hour session, Himmelfarb says. “Obviously nobody can carry that around them because it would weigh tons.” There are a few in-home models marketed as portable: Fresenius sells a device that it says gives patients more mobility. It weighs 34 kilograms and can be used with a home tap, as long as the water meets certain quality standards. But the first priority in making dialysis more convenient is to remove the need for an external water supply.

In Seattle, CDI researchers have developed a technique that pushes the used dialysis solution through a cartridge that uses light to convert urea – a key toxin targeted by dialysis – into nitrogen and carbon dioxide, so that the solution can be recycled³. The method can remove 15 grams of urea in 24 hours, sufficient for most people with kidney failure, and requires only 750 millilitres of solution, Himmelfarb says.

The team’s standalone haemodialysis device could be made compact enough to fit inside a rolling case, Himmelfarb says, weighing no more than 9 kilograms. Ideally, patients would use it daily, he says.

Another group trying to downsize dialysis was recently formed by the Dutch Kidney

rebalance the blood’s components and carry the toxins down the drain. Haemodialysis is much better at replicating the filtering than at achieving the subtle recalibration provided by the kidney tubules, says Jonathan Himmelfarb, the other co-director of the CDI. Healthy kidneys make subtle adjustments around the clock, whereas patients get just 12 hours of dialysis across 3 sessions each week, he says. Rebalancing the blood so abruptly can be a shock to the body and take hours to recover from. This is dubbed dialysis washout. Risher, who takes a shuttle to dialysis, often falls asleep on the way home.

The inefficient treatment is also very costly, as much as \$91,000 annually per patient in the United States. And as well as using a lot of water, the current approach consumes vast

quantities of power and materials such as plastics. “We use massive amounts of water – it’s not a green therapy,” Sedor says.

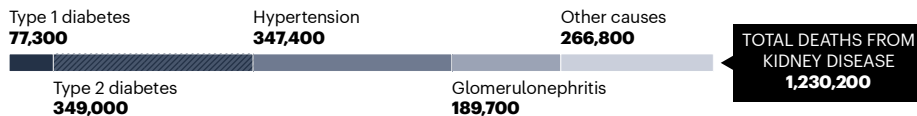
Plus, access is patchy around the globe. No more than one-third of people in Asia get dialysis, and even fewer in Africa¹ (see ‘A neglected need’). And even when patients in sub-Saharan Africa do begin the treatments, they are rarely able to sustain them for more than a few months².

The reason, in large part, is cost. Even when the government pays for the sessions, the patient’s family often has to foot the bill for lab tests, medications and other costs, says Gloria Ashuntantang, a nephrologist at Yaounde General Hospital in Cameroon. “Most of our patients will stop the therapy halfway, after having sold all property, and the children

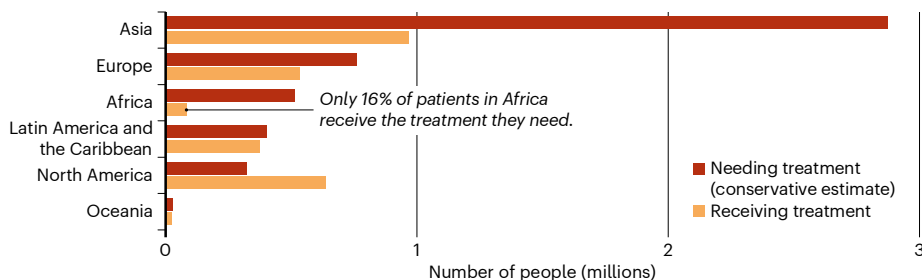
Feature

A NEGLECTED NEED

Chronic kidney disease kills more people each year than tuberculosis or HIV. Many cases are linked to hypertension and diabetes, which can be products of obesity, and a type of kidney inflammation called glomerulonephritis.



The number of people receiving treatment for kidney failure in the form of either dialysis or a transplant varies widely around the world.



Foundation, the medical-devices firm Debiotech in Lausanne, Switzerland, and non-profit insurers. Its latest prototype, which it hopes to make available to patients by 2023, weighs about 10 kilograms and will require only 6 litres of solution, according to Ton Rabelink, a nephrologist at Leiden University Medical Center in the Netherlands who is on the medical advisory board of the company, called NextKidney. The device, which could be used at home, limits the quantity of dialysis solution needed by using an absorbent material to soak up the toxins, Rabelink says.

In Singapore, researchers at the medical-technology company AWAK have been testing an even lighter device, one that weighs no more than 3 kilograms. It's designed for peritoneal dialysis, a technique that uses a catheter to send dialysis solution into the abdominal cavity, where a lining (the peritoneum) filters out toxins from the blood so they can drain, along with the solution, into an empty bag.

The AWAK device relies on a pump and a cartridge to absorb toxins from the used solution so that it can be recirculated. Each daily treatment would last seven to ten hours.

The company completed a safety trial involving 15 adults at Singapore General Hospital in 2018. It reported no serious adverse events, although some patients experienced abdominal discomfort or bloating. The device is one of several more-portable products in development that the FDA has agreed to expedite through its 'breakthrough devices' programme.

But testing a device in the controlled setting of a hospital is very different from using it in daily life, says Arshia Ghaffari, a researcher who directs dialysis services at the University of Southern California in Los Angeles. Furthermore, it's possible that the constant recirculation of dialysis solution could strain delicate membranes and "burn out the peritoneum faster", he says. A company spokesman discounted that concern, saying that the fluid is recirculated in small increments, just

250 millilitres at a time.

In some regions of the world, peritoneal dialysis is not an option, owing to the costs of shipping the heavy bags of solution. An international competition led by the George Institute for Global Health in Camperdown, Australia, in 2015 sought ways to improve access.

The winning technology, developed by Irish engineer Vincent Garvey, incorporates a lightweight kit that includes sterile bags contain-

THAT WOULD BE SO MUCH BETTER THAN ANY DEVICE, TO HAVE YOUR OWN KIDNEY FUNCTION REGENERATED."

ing a dry mix (dextrose and salts), along with a water distiller the size of a bread box, which sterilizes the water used to make the mix. A month's worth of supplies could be shipped in a box weighing 3 kilograms – a big improvement over a typical day's supply, which weighs 8 kilograms, says John Knight, managing director of Ellen Medical Devices in Camperdown, which was formed to develop the prototype. Knight's goal is to complete a clinical trial by the end of next year.

Recreating the kidney

Researchers at the University of California, San Francisco (UCSF), and Vanderbilt University in Nashville, Tennessee, have bypassed external devices and instead focused on developing a kidney prototype that they hope will one day be surgically implanted into a patient's body. It wouldn't require a pump because it would

be attached to key arteries and powered by blood pressure, says Vanderbilt nephrologist William Fissell, who co-directs the research with UCSF's Shuvo Roy.

The device contains two key parts: a blood-filtration system and a cell-infused recalibration module. The filter is made of silicon membranes with nanometre-scale pores that are designed to mimic the glomerulus. The recalibration module uses tubule cells from discarded human kidneys to rebalance the blood's components, Fissell says.

Late last year, researchers reported at an American Society of Nephrology meeting that they had conducted the first safety test of the recalibration module in pigs, without any of the serious problems often seen with implanted devices, including an immune reaction or blood clots.

But Rabelink thinks that implantable devices will be more difficult to develop, given that they rely on a mix of engineered and biological elements, complicating the design and creating extra regulatory hurdles. In the meantime, he posits that advances in stem-cell research might surpass such efforts. "In the end, that would be so much better than any device, to have your own kidney function regenerated or prolonged," he says.

But Fissell and Roy counter that stem-cell techniques have been slow to pay off in other areas, such as diabetes treatment, so devices such as automated insulin pumps have led the way. Fissell describes the project's primary hurdle as securing sufficient funding to manufacture the device, which is roughly the size of a soft-drink can, on a larger, standardized scale so it can be evaluated by US regulators. "I've got the [device] right on my desk – it's ready to sew," he says.

Despite the confidence of some teams, Sheldon thinks that recreating the sophistication of a kidney is too complex for any single team, and will probably require a mix of engineering and biology, plus a lot more money. He proposed the idea of an international coalition at an American Society of Nephrology meeting last year, and has planned a series of meetings in Europe later this year with stakeholders and medical groups.

For Risher and other patients, access to any portable device would be liberating, providing "that freedom and flexibility to do dialysis when I want to do it", he says. As a car aficionado, he dreams of throwing his machine onto the passenger seat and steering for the open road, only the horizon before him.

Charlotte Huff is a science journalist based in Fort Worth, Texas.

1. Liyanage et al. *Lancet* **385**, 1975–1982 (2015).
2. Ashuntantang, G. et al. *Lancet Glob. Health* **5**, e408–17 (2017).
3. Shao, G., Zang, Y. & Hinds, B. *Appl. Nano Mater.* **2**, 6116–6123 (2019).

SOURCES: MORTALITY: GBD CHRONIC KIDNEY DISEASE COLLABORATION; LANCET 395, 709–733 (2020); TREATMENT: REF. 1

Comment



ILLUSTRATION BY DAVID PARKINS

A controlled trial for reproducibility

Marc P. Raphael, Paul E. Sheehan & Gary J. Vora

For three years, part of DARPA has funded two teams for each project: one for research and one for reproducibility. The investment is paying off.

In 2016, the US Defense Advanced Research Projects Agency (DARPA) told eight research groups that their proposals had made it through the review gauntlet and would soon get a few million dollars from its Biological Technologies Office (BTO). Along with congratulations, the teams received a reminder that their award came with an unusual requirement – an independent shadow team of scientists tasked with reproducing their results.

Thus began an intense, multi-year controlled trial in reproducibility. Each shadow team consists of three to five researchers, who visit the ‘performer’ team’s laboratory and often host visits themselves. Between 3% and 8% of the programme’s total funds go to this independent validation and verification (IV&V) work. But DARPA has the flexibility and

resources for such Herculean efforts to assess essential techniques. In one unusual instance, an IV&V laboratory needed a sophisticated US\$200,000 microscopy and microfluidic set-up to make an accurate assessment.

These costs are high, but we think they are an essential investment to avoid wasting taxpayers’ money and to advance fundamental research towards beneficial applications. Here, we outline what we’ve learnt from implementing this programme, and how it could be applied more broadly.

Engineering lessons

Engineers expect their work to be subject to an IV&V process, in which the organization conducting the research uses a separate set of engineers to test, for example, whether microprocessors or navigation software work as expected. NASA’s IV&V facility was established more than 25 years ago and has around 300 employees testing code and satellite components.

In conventional electronics, IV&V relies on fundamental units such as transistors, diodes, capacitors and oscillators. The electronics industry takes great pains to promote the compatibility of these basic elements across platforms. For example, a given microprocessor developed by Intel can function in both

children's toys and advanced physics labs.

That compatibility does not apply to biological systems: proteins and cell lines are often not interchangeable, and even identical proteins can function differently in different environments. First-principle explanations of how they work often do not exist. This complicates efforts to apply IV&V approaches developed for electronics and software. What's more, increasingly sophisticated bioengineering tools are making cell-biology experiments more complicated, so thorough validation could take months or even years to complete. Few investigators have such resources.

Instead, the biological sciences have depended on other, less-reliable techniques for reproducibility. The most long-standing is the assumption that reproducibility studies will occur organically as different researchers work on related problems. In the past five years or so, funding agencies and journals have implemented more-stringent experimental-reporting and data-availability requirements for grant proposals and submitted manuscripts. A handful of initiatives have attempted to replicate select published studies. The peer-reviewed *Journal of Visualized Experiments* creates videos to disseminate details that are hard to convey in conventional methods sections.

Yet pitfalls persist. Scientists might waste resources trying to build on unproven techniques. And real discoveries can be labelled irreproducible because too few resources are available to conduct a validation. We were lucky enough to have the time, money and mandate to try something different.

Making it work

The synthetic-biology focus of DARPA's Biological Control programme is well suited to merging biological research with reproducibility studies. The programme aims to bring engineering principles of design and control to biology. By definition, this requires the adoption of best practices from the engineering community – such as IV&V – to improve the likelihood that technologies can advance.

Awardees were told from the outset that they would be paired with an IV&V team consisting of unbiased, third-party scientists hired by and accountable to DARPA. In this programme, we relied on US Department of Defense laboratories, with specific teams selected for their technical competence and ability to solve problems creatively. To get comfortable with the concept of IV&V, investigators needed reassurance that replicating teams would not steal ideas or derail

publications. They also needed to get used to their results being challenged even before peer-review submission, and they needed reminders that cooperating with these teams was a programme requirement.

Results so far show a high degree of experimental reproducibility. The technologies investigated include using chemical triggers to control how cells migrate¹; introducing synthetic circuits that control other cell functions²; intricate protein switches that can be programmed to respond to various cellular conditions³; and timed bacterial expression that works even in the variable environment of the mammalian gut⁴. In the future, we expect replication efforts will be reported as supplemental data submitted with manuscripts.

“Real discoveries can be labelled irreproducible because too few resources are available to conduct a validation.”

Especially when claims border on the fantastical, it is helpful to show peer reviewers and editors that an independent party has confirmed the finding. So far, one publication co-authored by performer and IV&V teams has been accepted⁵, and two more are nearing submission. Still, getting to this point was more difficult than we expected. It demanded intense coordination, communication and attention to detail.

Successfully combining reproducibility studies with fundamental research required a level of coordination between laboratories and with the programme manager (P.E.S.) that none of us had experienced before. The manager worked with each project team to determine which of their many results merited validation on the basis of the desired impact and application. We wanted to know that the engineered organism – yeast, bacteria, slime moulds, mammalian cells or something else – could be modified reliably and that these modifications performed as expected, as well as what environmental conditions were essential for that performance.

A typical academic lab trying to reproduce another lab's results would probably limit itself to a month or so and perhaps three or four permutations before giving up. Our effort needed capable research groups that could dedicate much more time (in one case, 20 months) and that could flexibly follow evolving research.

Ultimately, the technologies that DARPA

is developing should end up being applied by many people for a broad range of uses. So in addition to assessing whether the technologies worked, IV&V teams had to assess robustness. For instance, we needed to know what fraction of cells would incorporate new genetic material, especially when multiple genes and control elements were involved. We tested whether cells would still work in the same way if frozen and thawed months later, and whether they would retain their functionality after being grown continuously. One IV&V team checked whether migration in a genetically modified cell line was faster than in its precursor, and fabricated guidance chips to determine what surfaces best directed cell migration.

Achieving verification means communicating effectively. Performer teams, particularly those with several principal investigators, had to designate someone to facilitate teleconferences and site visits. Both teams present jointly to the programme manager at least twice a year.

A key component of the IV&V teams' effort has been to spend a day or more working with the performer teams in their laboratories. Often, members of a performer laboratory travel to the IV&V laboratory as well. These interactions lead to a better grasp of methodology than reading a paper, frequently revealing person-to-person differences that can affect results. This is especially true when the IV&V investigator does not regularly work with the same cell type as the performer team, and thus approaches experiments in a similar way to other researchers who are building on a newly reported technique.

Real-time collaboration minimizes or avoids logistical roadblocks that are known to prevent basic research validation (for example, when the original samples cannot be located, or the postdoctoral researcher with the necessary expertise is no longer with the laboratory). Still, our IV&V efforts have been derailed for weeks at a time for trivial reasons (see 'Hard lessons'), such as a typo that meant an ingredient in cell media was off by an order of magnitude. We lost more than a year after discovering that commonly used biochemicals that were thought to be interchangeable are not. A five-laboratory consortium testing how cultured cells responded to cancer drugs reported similar experiences, with minor differences causing major effects⁶.

Now, our IV&V efforts begin by cataloguing all chemicals, media and cell types, their suppliers and, for animal-derived extracts, lot numbers. Instruments are calibrated

HARD LESSONS

Recommendation	What to do	Our experience
Document reagents	Include the vendor, product number and lot number for all reagents.	We lost weeks of work and performed useless experiments when we assumed that identically named reagents (for example, polyethylene glycol or fetal bovine serum) from different vendors could be used interchangeably.
See it live	Watch an experiment carried out by another team. In our hands, washing cells too vigorously or using the wrong-size pipette tip changed results unpredictably.	Site visits are mandatory because witnessing experiments in action reveals valuable information, such as how to trap <i>Hydra</i> without harming them, or how to tilt a cell plate. The benefits of site visits in terms of achieving reproducibility are worth the cost of plane tickets and lodging.
State a range	Rather than a single number, state a range of acceptable conditions for temperature, convection and other control standards.	Knowing whether 21 °C means 20.5–21.5 °C or 20–22 °C can tell you whether cells will thrive or wither, and whether you'll need to buy an incubator to make an experiment work.
Test, then ship	Immediately before shipping cells or a genetic construct for testing, check them or it.	Incorrect, outdated or otherwise diminished products were sent to the IV&V team for verification many times.
Double check	If a standard protocol does not work, the performer and independent valuation and verification (IV&V) teams should work together on a step-by-step review.	A typo in one protocol cost us four weeks of failed experiments, and in general, vague descriptions of formulation protocols (for example, for expressing genes and making proteins without cells) caused months of delay and cost thousands of dollars in wasted reagents.
Pick a person	Each performer team should designate one person to keep communication open, accurate and timely.	The projects that lacked a dedicated and stable point of contact were the same ones that took the longest to reproduce. That is not coincidence.
Keep <i>in silico</i> analysis up to date	Data-analysis pipelines are replete with configuration decisions, assumptions, dependencies and contingencies that move quickly beyond documentation, making troubleshooting incredibly difficult.	Teams had to visit each others' labs more than once to understand and fully implement computational-analysis pipelines for large microscopy data sets.

and settings (for example, microscope light source power) coordinated to avoid unintended side effects such as phototoxicity. Routine measurements in cell culture, such as pH, osmolarity and testing for *Mycoplasma*, which often fall by the way-side, are prioritized. Each project creates a customized checklist depending on its cell lines, equipment and experiments. Without this essential level of research hygiene, troubleshooting efforts would become an uninformative time sink.

We have learnt to note the flow rates used when washing cells from culture dishes, to optimize salt concentration in each batch of medium and to describe temperature and other conditions with a range rather than a single number. This last practice came about after we realized that diminished slime-mould viability in our Washington DC facility was due to lab temperatures that could fluctuate by 2 °C on warm summer days, versus the more tightly controlled temperature of the performer lab in Baltimore 63 kilometres away. Such observations can be written up in a protocol paper.

Sometimes, validation requires new equipment. For the slime moulds, independent validation meant buying an incubator that could keep cells stably at 21.5 °C, slightly below the IV&V laboratory's ambient temperature. In another case, the performer team had to help install customized microfluidic and optical equipment at the IV&V lab because the standard microscopes and analysis software used for live-cell imaging were not up to the task.

All this makes for a considerably more variable IV&V programme than is found in microelectronics. But without these efforts, some promising technologies could have

been abandoned prematurely as seeming dead ends.

Big dividends

We think that the IV&V programme brings benefits beyond reproducing any individual project. Now, there is a process to make investigations of disparate results more transparent. Performing reproducibility studies invariably forces scientists to think more deeply about their own experimental protocols and techniques. As one of our scientists said, "IV&V forces performers to think more critically about what qualifies as a successful system, and facilitates candid discussion about system performance and limitations." Trainees told us that they have gained skill in analysing data, providing constructive criticism and designing and documenting their own research so that it can be reproduced.

IV&V teams gained further advantages. For example, because service laboratories become well-versed in the mindset and protocols for new technologies even before publications appear, they are well-poised to integrate them into their offerings, predict future directions for the field and move research more quickly to applications. The IV&V programme also expands networking opportunities between DARPA scientists and the top-quality labs DARPA funds, including the potential to recruit postdocs and graduate students across laboratories. Not surprisingly, many DARPA BTO programmes in recent years have incorporated some form of IV&V to help validate programme results.

As we continue the Biological Control IV&V programme, we expect to find more ways to improve it, to better quantify its benefits and to codify best practices, such as incorporating

automation and robotics where possible and keeping an open line of communication between performer groups and IV&V teams. Although some of the lessons learnt from the first stages might seem obvious and trite, that also reinforces their necessity.

We think that a dedicated shift towards the IV&V model by more research institutions and funding agencies will bring more reliable and cost-effective science. Programme officers at other granting agencies should consider allocating a portion of their funding stream to independent reproducibility efforts. This will both reduce the number of papers that cannot be replicated and improve the quality of work that funding agencies support. Metrics will need to be established to quantify the cost savings of applying this model to synthetic biology and bioengineering, but given its successful integration throughout more conventional engineering disciplines, we are optimistic that the returns will be worth it.

The authors

Marc P. Raphael is a biophysicist at the Naval Research Laboratory in Washington DC, USA.

Paul E. Sheehan is a programme manager in DARPA's Biological Technologies Office in Arlington, Virginia, USA. **Gary J. Vora** is a biologist at the Naval Research Laboratory in Washington DC, USA.

e-mail: marc.raaphael@nrl.navy.mil

1. Miao, Y. C. et al. *Nature Cell Biol.* **19**, 329–340 (2017).
2. Ng, A. H. et al. *Nature* **572**, 265–269 (2019).
3. Langan, R. A. et al. *Nature* **572**, 205–210 (2019).
4. Riglar, D. T. et al. *Nature Commun.* **10**, 4665 (2019).
5. Ames, C. L. et al. *Commun. Biol.* **3**, 67 (2020).
6. Niepel, M. et al. *Cell Syst.* **9**, 35–48 (2019).

Correspondence

COVID-19: don't ignore Taiwan

As the COVID-19 epidemic unfolds, history is repeating itself in Taiwan. Still denied membership of the World Health Organization (WHO) and therefore participation in international decisions, Taiwan could again experience a disproportionate number of deaths, as it did during the 2003 SARS epidemic (see *Nature* **422**, 652; 2003). Taiwan needs help and, in turn, the WHO has everything to gain from allowing it to join in the fight against this crisis.

The COVID-19 epidemic calls for a response consistent with the principles of the WHO (see go.nature.com/2tbgqrd). In my view, Taiwan's alienation is an inexcusable liability for global health. Its health-care system is ranked first in the world by NUMBEO (see go.nature.com/2wbqckc). Its researchers identified receptor-binding proteins of the 2003 SARS virus (see go.nature.com/3cqqn82), established animal models for testing vaccines against it, and are now pursuing vaccine research and development against COVID-19.

Taiwan is separated from mainland China by a mere strait, across which thousands travelled to Taiwan every day until Taiwan imposed entry restrictions last month because of the epidemic. The WHO should look again at its exclusion of Taiwan. There is no place for political disputes when millions of lives are at stake.

Chase W. Nelson Biodiversity Research Center, Academia Sinica, Taipei, Taiwan.
cnelson@gate.sinica.edu.tw

COVID-19: keep up with latest papers

An open-resource literature hub known as LitCovid curates the most comprehensive collection of international research papers so far on the new coronavirus disease COVID-19 (see go.nature.com/3almd5p). Developed with the support of the US National Institutes of Health's intramural research programme, LitCovid is updated daily with newly published articles. The aim is to provide timely insight from the scientific literature into the biology of the virus and the diagnosis and management of those who have been infected.

LitCovid has a more sophisticated search function than existing resources. It identifies roughly 35% more relevant articles than do conventional keyword-based searches for entries such as 'COVID-19' or 'nCoV'. Furthermore, the articles are categorized by topic – overview, disease mechanism, transmission dynamics, treatment, case report and epidemic forecasting – as well as by geographic location for visualization on a world map.

We welcome user feedback for further enhancement.

Qingyu Chen, Alexis Allot, Zhiyong Lu National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA.
zhiyong.lu@nih.gov

Disaster research: for volunteers only

In our view, a code of conduct for research done in disaster zones should include a guarantee that people in the area can choose whether or not to participate (see J. C. Gaillard and L. Peek *Nature* **575**, 440–442; 2019). Fair and voluntary participation is a fundamental human right.

Residents in a disaster area fear for their lives in the acute phase, and face health risks in the recovery phase. They naturally seek help and support from their government and from professionals. Under such circumstances, residents might agree to take part in research without giving the matter enough thought. It is therefore important for researchers to make clear to them that studies could have harmful effects as well as benefits.

As an example, after the 2011 accident at the Fukushima Daiichi nuclear-power station, we undertook screening of local children for thyroid cancer, which can be induced by radioactive iodine (A. Ohtsuru *et al. JAMA Otolaryngol. Head Neck Surg.* **145**, 4–11; 2019). Although participants and their parents gave written consent, they were unaware of the risks of overdiagnosis, which include having unnecessary surgery to remove small, slow-growing tumours (see go.nature.com/2vfd9z7). Also, screening conducted during school time could have given the impression that participation was mandatory.

Such ambiguity underlines the importance of a code of conduct that makes participation in disaster-zone research voluntary.

Sanae Midorikawa, Akira Ohtsuru Fukushima Medical University, Fukushima, Japan.
otobokesoh@gmail.com

Disaster research: feedback to society

In our experience, the recovery of disaster zones can be improved by providing research feedback to affected communities (see J. C. Gaillard and L. Peek *Nature* **575**, 440–442; 2019).

In 2017, we questioned communities that had been evacuated following the 2011 Tohoku earthquake and Fukushima nuclear accident in Japan. We asked them about their awareness of the risks of living in the region and the factors that had been important for recovery. We then fed our results back to them – much to their surprise, because we were the first researchers ever to have done so. These communities knew nothing about the 20,000 publications related to the disaster.

In our view, this remoteness of researchers from those affected by such catastrophes needs to be corrected. It could stem from mere oversight or from a reluctance to sacrifice time that might otherwise be spent writing papers or grant applications. However, meetings with local communities before and after data collection are ethical and productive. Furthermore, reporting the details to guide the next generation of disaster-zone researchers should be mandatory.

Tomoyuki Kobayashi, Yoshitake Takebayashi, Michio Murakami Fukushima Medical University, Fukushima, Japan.
tomokoba@fmu.ac.jp

News & views

Geochemistry

A hint of Earth's ancient ingredients

Katherine R. Bermingham

Identifying Earth's building blocks from terrestrial rocks is challenging because these ingredients have become mixed as the planet evolved. Evidence of an unknown building block in ancient rocks provides fresh insight. **See p.240**

Earth formed from an unknown selection of meteoritic material. On page 240, Fischer-Gödde *et al.*¹ report that the composition of ruthenium isotopes in ancient rocks from southwest Greenland contains evidence of a previously unrecognized building block of Earth. Surprisingly, the inferred isotopic composition of ruthenium in the material does not match known meteorite compositions. The authors' findings suggest that Earth's volatile components, such as water and organic compounds, could have arrived during the final stages of the planet's growth.

Our planet is the product of a series of collisions of increasingly large celestial bodies^{2–4}. These building blocks accreted from a protoplanetary disk of dust and gas that orbited the proto-Sun about 4.6 billion years ago. Identifying the compositions of Earth's building blocks is difficult because of our limited access to the disk's remnants, and because of the complex, long-term geological processing of the mantle that has mixed Earth's ancient ingredients.

Potential answers to the question of what Earth is made of can come from studies in which the isotopic compositions of terrestrial rock samples are compared with those of meteorites that formed within the first few million years of the Solar System's history. These meteorites are presumed to be representative of the smaller bodies that ultimately coalesced to form the rocky planets. Consequently, meteorites are our most promising candidates for Earth's building blocks.

Fischer-Gödde and colleagues' study builds on the finding that meteorites have characteristic isotope compositions that serve as fingerprints to distinguish different types of potential building block. For example, meteorites such as carbonaceous chondrites,

which are often 'wet' (that is, they contain volatile components), have different isotopic fingerprints from meteorites that are generally 'dry'⁵. The differences in isotopic composition originate from the heterogeneous distribution of stardust in the protoplanetary disk, and are known as nucleosynthetic isotope variations. If the fingerprints could be identified in terrestrial rock samples, this might provide evidence of the material from meteorites that Earth was built from.

The documentation of fingerprints in terrestrial rocks could help to constrain estimates of when volatile elements were delivered to Earth and where they came from. This is because the abundances of certain isotopes of some elements – ruthenium-100 (¹⁰⁰Ru),

for example – not only distinguish between wet and dry building blocks, but also trace different stages of Earth's accretion history.

Ruthenium is classified as a highly siderophile (iron-loving) element, because it collects in metal-rich phases of Earth's interior. Consequently, most of our planet's ruthenium is concentrated in its metallic core. There are, however, traces of ruthenium and other highly siderophile elements (HSEs) in the mantle, and their relative proportions approximate to those measured in primitive meteorites⁶. One interpretation of this is that the HSEs were added to the mantle after the core formed, during an event called the late veneer – when the final approximately 0.5% (of the total percentage weight) of Earth's mass accreted^{7,8}. If so, then ruthenium and other HSEs in the mantle record the composition of the last material that accreted to Earth⁹.

It has been proposed that Earth's volatile elements were also added during the late veneer, possibly by the accretion of carbonaceous chondrites^{10,11}. Studies in the past few years, however, have found a mismatch between the ¹⁰⁰Ru-isotope composition (the abundances of ¹⁰⁰Ru in terrestrial rocks) in Earth's mantle and that in carbonaceous chondrites^{12,13}. It could therefore be concluded that carbonaceous chondrites did not form part of the late veneer, thus casting doubt on the timing of the delivery of volatiles to Earth¹³.

This conclusion rests on the assumption that HSEs in the mantle do not contain significant quantities of material from before the late

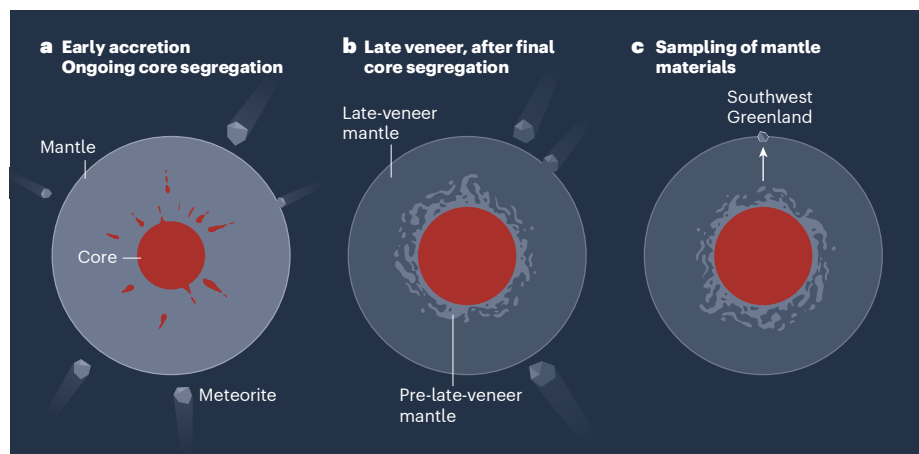


Figure 1 | A scenario for the preservation of ancient material in Earth's mantle. a, Between 4.6 billion and about 4.5 billion years ago, Earth formed from the accretion of material from meteorites. Siderophile elements, which have a strong affinity for metals, segregated into the core. **b**, The final approximately 0.5% of the total percentage weight of Earth's mass accreted from meteorites during an event called the late veneer, after the core had formed. **c**, Fischer-Gödde *et al.*¹ report that ancient rocks from southwest Greenland have an unusual ruthenium-isotope composition. They attribute this to the presence of pre-late-veneer mantle material in the rocks. The distribution of pre-late-veneer material shown here is speculative; the actual amount and distribution cannot be derived from the available data.

veneer – a reasonable assertion, given that there is limited direct evidence of this. If the pre-late-veneer mantle did contain a substantial amount of ^{100}Ru that did not collect in the core, and that was identifiable by having a different ^{100}Ru -isotope composition from that of the modern mantle, then carbonaceous chondrites could still have been accreted during the late veneer.

Nucleosynthetic ruthenium-isotope variations have not been reported for terrestrial rocks before now. This is, in part, because Earth has active plate tectonics and mantle convection, which mix and dilute the fingerprints of its building blocks. However, in the past few years, analytical methods¹⁴ have been further developed that enable isotope variations to be measured on the scale of parts per million, making it possible to search for these primitive isotopic signatures.

By comparing the ^{100}Ru -isotope compositions of terrestrial rocks with those of meteorites, Fischer-Gödde and co-workers report that an ancient part of Earth, preserved in rocks from southwest Greenland, retains the fingerprints of an unusual building block (Fig. 1). The fact that the inferred isotope compositions do not match known meteorite compositions indicates that current meteorite collections are considerably limited in their sampling of the protoplanetary disk.

The authors interpret their unusual ^{100}Ru data as the isotopic signature of pre-late-veneer ruthenium in the source of these rocks. Considering their findings in the context of the compositions of other HSEs in the mantle, the authors suggest that the modern composition of the mantle can be reconciled with their new data only if the late veneer contained carbonaceous chondrites to counterbalance the composition of the pre-late-veneer component of the mantle. This would mean that volatiles could have been delivered to Earth during the final stages of the planet's formation.

Fischer-Gödde and colleagues' data answer the long-standing question of whether Earth's diverse building blocks are preserved and accessible for study. But the data also raise key questions, the answers to which will undoubtedly determine the importance of the new findings. For example, how representative of the pre-late-veneer mantle is the suite of rock samples from southwest Greenland? Are nucleosynthetic fingerprints observed in the isotopic compositions of other elements in the mantle? What is the composition of the 'missing' meteorites that dominated the ruthenium composition of the pre-late-veneer mantle, and why has it not yet been identified? And how was the isotopic signature of these meteorites preserved in the convecting mantle? These questions can be addressed only by expanding the search for nucleosynthetic fingerprints in the mantle.

Katherine R. Bermingham is in the Department of Earth and Planetary Sciences, Rutgers University, Busch Campus, Piscataway, New Jersey 08854, USA.
e-mail: katherine.bermingham@rutgers.edu

1. Fischer-Gödde, M. *et al.* *Nature* **579**, 240–244 (2020).
2. Chambers, J. E. *Astrophys. J.* **825**, 63–81 (2016).
3. Lambrechts, M. & Johansen, A. *Astron. Astrophys.* **544**, A32 (2012).
4. Morbidelli, A. & Nesvorný, D. *Astron. Astrophys.* **546**, A18 (2012).
5. Warren, P. H. *Earth Planet. Sci. Lett.* **311**, 93–100 (2011).

6. Becker, H. *et al.* *Geochim. Cosmochim. Acta* **70**, 4528–4550 (2006).
7. Chou, C.-L. *Proc. 9th Lunar Planet. Sci. Conf.* 219–230 (Lunar Planet. Inst., 1978).
8. Kimura, K., Lewis, R. S. & Anders, E. *Geochim. Cosmochim. Acta* **38**, 683–781 (1974).
9. Dauphas, N. *Nature* **541**, 521–524 (2017).
10. Marty, B. *Earth Planet. Sci. Lett.* **313–314**, 56–66 (2012).
11. Alexander, C. M. O'D. *et al.* *Science* **337**, 721–723 (2012).
12. Bermingham, K. R. & Walker, R. J. *Earth Planet. Sci. Lett.* **474**, 466–473 (2017).
13. Fischer-Gödde, M. & Kleine, T. *Nature* **541**, 525–527 (2017).
14. Fischer-Gödde, M., Burkhardt, C., Kruijer, T. S. & Kleine, T. *Geochim. Cosmochim. Acta* **168**, 151–171 (2015).

Tumour biology

Tweaking DNA of myeloid cells curbs cancer spread

Ali Ghasemi & Michele De Palma

Blood cells called myeloid cells can facilitate metastasis – the spread of a tumour to distant organs. Taming these cells with drugs that alter the chemical structure of their DNA limits metastasis in mice. **See p.284**

Many tumours are successfully treated with surgery. But cancer might recur at the surgical site or in distant organs, so surgery is often followed by treatment termed adjuvant therapy. This limits the risk of relapse by killing cancer cells remaining at the surgical site or those that have already moved elsewhere. However, adjuvant therapy is not always effective. Moreover, it might not prevent certain processes that aid cancer resurgence, such as the recruitment of blood cells called myeloid cells to distant organs, where they can lay the foundation for cancer cells to settle and thrive¹. *Lu et al.*² reveal on page 284 how the chemical structure of the DNA in the nucleus of myeloid cells is a vulnerability that can be harnessed to target these tumour-promoting cells and limit cancer spread.

Tumour spread from its primary site to distant organs, which is called metastasis, involves complex interactions between cancer cells and the surrounding healthy tissues. Evidence is growing that primary tumours can produce signals that modify normal cells to generate a 'soil' in distant organs – termed a pre-metastatic niche – that permits subsequent 'seeding' and establishment of cancer cells at this secondary site³. Such secondary tumours, or metastases, are often lethal.

Several types of cancer metastasize from their primary site to the lung. Efficient metastasis of breast cancer cells to the lung in mice requires the participation of a type of myeloid cell called a monocyte^{4,5}. These cells normally function to fight infections, but they

can also supply metastasizing cancer cells with factors that help them to get established and grow at a secondary site^{1,4}. Certain molecules produced by the primary tumour alter the properties of monocytes and increase their numbers in the bloodstream, fostering the monocytes' tumour-supporting functions in the pre-metastatic niche^{1,4}. Therefore, blocking such 'tumour-educated' monocytes might inhibit metastasis. Previous work^{1,4} indicates that neutralizing a protein called CCL2, which is produced by the tumour and promotes monocyte accumulation in the lungs of mice, impairs metastasis in mouse tumour models. However, this approach was unsuccessful in clinical trials owing to difficulties in effectively neutralizing CCL2 (ref. 4).

Lu et al. studied mice that were given transplants under the skin of a type of tumour that metastasizes to the lung. Consistent with previous studies^{4,5}, the authors found that two types of myeloid cell – monocytes and neutrophils – accumulated in the lung before metastases were detectable there. These tumour-elicited myeloid cells are collectively called myeloid-derived suppressor cells (MDSCs) owing to their ability to suppress the immune response against a tumour⁵. Confirming the metastasis-promoting capacity of MDSCs, the authors report that elimination of MDSCs delayed the metastasis of cancer cells to the lung and extended the animals' survival.

Gene expression can be regulated by changes in the nucleus termed epigenetic

veneer – a reasonable assertion, given that there is limited direct evidence of this. If the pre-late-veneer mantle did contain a substantial amount of ^{100}Ru that did not collect in the core, and that was identifiable by having a different ^{100}Ru -isotope composition from that of the modern mantle, then carbonaceous chondrites could still have been accreted during the late veneer.

Nucleosynthetic ruthenium-isotope variations have not been reported for terrestrial rocks before now. This is, in part, because Earth has active plate tectonics and mantle convection, which mix and dilute the fingerprints of its building blocks. However, in the past few years, analytical methods¹⁴ have been further developed that enable isotope variations to be measured on the scale of parts per million, making it possible to search for these primitive isotopic signatures.

By comparing the ^{100}Ru -isotope compositions of terrestrial rocks with those of meteorites, Fischer-Gödde and co-workers report that an ancient part of Earth, preserved in rocks from southwest Greenland, retains the fingerprints of an unusual building block (Fig. 1). The fact that the inferred isotope compositions do not match known meteorite compositions indicates that current meteorite collections are considerably limited in their sampling of the protoplanetary disk.

The authors interpret their unusual ^{100}Ru data as the isotopic signature of pre-late-veneer ruthenium in the source of these rocks. Considering their findings in the context of the compositions of other HSEs in the mantle, the authors suggest that the modern composition of the mantle can be reconciled with their new data only if the late veneer contained carbonaceous chondrites to counterbalance the composition of the pre-late-veneer component of the mantle. This would mean that volatiles could have been delivered to Earth during the final stages of the planet's formation.

Fischer-Gödde and colleagues' data answer the long-standing question of whether Earth's diverse building blocks are preserved and accessible for study. But the data also raise key questions, the answers to which will undoubtedly determine the importance of the new findings. For example, how representative of the pre-late-veneer mantle is the suite of rock samples from southwest Greenland? Are nucleosynthetic fingerprints observed in the isotopic compositions of other elements in the mantle? What is the composition of the 'missing' meteorites that dominated the ruthenium composition of the pre-late-veneer mantle, and why has it not yet been identified? And how was the isotopic signature of these meteorites preserved in the convecting mantle? These questions can be addressed only by expanding the search for nucleosynthetic fingerprints in the mantle.

Katherine R. Bermingham is in the Department of Earth and Planetary Sciences, Rutgers University, Busch Campus, Piscataway, New Jersey 08854, USA.
e-mail: katherine.bermingham@rutgers.edu

1. Fischer-Gödde, M. *et al.* *Nature* **579**, 240–244 (2020).
2. Chambers, J. E. *Astrophys. J.* **825**, 63–81 (2016).
3. Lambrechts, M. & Johansen, A. *Astron. Astrophys.* **544**, A32 (2012).
4. Morbidelli, A. & Nesvorný, D. *Astron. Astrophys.* **546**, A18 (2012).
5. Warren, P. H. *Earth Planet. Sci. Lett.* **311**, 93–100 (2011).

6. Becker, H. *et al.* *Geochim. Cosmochim. Acta* **70**, 4528–4550 (2006).
7. Chou, C.-L. *Proc. 9th Lunar Planet. Sci. Conf.* 219–230 (Lunar Planet. Inst., 1978).
8. Kimura, K., Lewis, R. S. & Anders, E. *Geochim. Cosmochim. Acta* **38**, 683–781 (1974).
9. Dauphas, N. *Nature* **541**, 521–524 (2017).
10. Marty, B. *Earth Planet. Sci. Lett.* **313–314**, 56–66 (2012).
11. Alexander, C. M. O'D. *et al.* *Science* **337**, 721–723 (2012).
12. Bermingham, K. R. & Walker, R. J. *Earth Planet. Sci. Lett.* **474**, 466–473 (2017).
13. Fischer-Gödde, M. & Kleine, T. *Nature* **541**, 525–527 (2017).
14. Fischer-Gödde, M., Burkhardt, C., Kruijer, T. S. & Kleine, T. *Geochim. Cosmochim. Acta* **168**, 151–171 (2015).

Tumour biology

Tweaking DNA of myeloid cells curbs cancer spread

Ali Ghasemi & Michele De Palma

Blood cells called myeloid cells can facilitate metastasis – the spread of a tumour to distant organs. Taming these cells with drugs that alter the chemical structure of their DNA limits metastasis in mice. **See p.284**

Many tumours are successfully treated with surgery. But cancer might recur at the surgical site or in distant organs, so surgery is often followed by treatment termed adjuvant therapy. This limits the risk of relapse by killing cancer cells remaining at the surgical site or those that have already moved elsewhere. However, adjuvant therapy is not always effective. Moreover, it might not prevent certain processes that aid cancer resurgence, such as the recruitment of blood cells called myeloid cells to distant organs, where they can lay the foundation for cancer cells to settle and thrive¹. *Lu et al.*² reveal on page 284 how the chemical structure of the DNA in the nucleus of myeloid cells is a vulnerability that can be harnessed to target these tumour-promoting cells and limit cancer spread.

Tumour spread from its primary site to distant organs, which is called metastasis, involves complex interactions between cancer cells and the surrounding healthy tissues. Evidence is growing that primary tumours can produce signals that modify normal cells to generate a 'soil' in distant organs – termed a pre-metastatic niche – that permits subsequent 'seeding' and establishment of cancer cells at this secondary site³. Such secondary tumours, or metastases, are often lethal.

Several types of cancer metastasize from their primary site to the lung. Efficient metastasis of breast cancer cells to the lung in mice requires the participation of a type of myeloid cell called a monocyte^{4,5}. These cells normally function to fight infections, but they

can also supply metastasizing cancer cells with factors that help them to get established and grow at a secondary site^{1,4}. Certain molecules produced by the primary tumour alter the properties of monocytes and increase their numbers in the bloodstream, fostering the monocytes' tumour-supporting functions in the pre-metastatic niche^{1,4}. Therefore, blocking such 'tumour-educated' monocytes might inhibit metastasis. Previous work^{1,4} indicates that neutralizing a protein called CCL2, which is produced by the tumour and promotes monocyte accumulation in the lungs of mice, impairs metastasis in mouse tumour models. However, this approach was unsuccessful in clinical trials owing to difficulties in effectively neutralizing CCL2 (ref. 4).

Lu et al. studied mice that were given transplants under the skin of a type of tumour that metastasizes to the lung. Consistent with previous studies^{4,5}, the authors found that two types of myeloid cell – monocytes and neutrophils – accumulated in the lung before metastases were detectable there. These tumour-elicited myeloid cells are collectively called myeloid-derived suppressor cells (MDSCs) owing to their ability to suppress the immune response against a tumour⁵. Confirming the metastasis-promoting capacity of MDSCs, the authors report that elimination of MDSCs delayed the metastasis of cancer cells to the lung and extended the animals' survival.

Gene expression can be regulated by changes in the nucleus termed epigenetic

modifications. These include processes that add or remove either methyl groups on DNA or acetyl groups on DNA-binding proteins called histones. Such epigenetic modifications often alter the expression of many genes in cancer cells, which might lead to loss of expression of genes that function as tumour suppressors, or to high expression of tumour-promoting genes. Anticancer treatments that use drugs to modulate epigenetic changes (termed epigenetic therapies) are gaining momentum⁶. Some of these drugs have been tested in clinical trials and approved for the treatment of certain blood tumours⁶.

Epigenetic therapy is usually aimed at cancer cells, but earlier work^{7,8} indicates that such treatment can also target tumour-promoting myeloid cells. Lu and colleagues investigated the effect of using low doses of two such drugs: 5-azacytidine (which inhibits the addition of methyl groups to DNA) and entinostat (which inhibits the removal of acetyl groups from histones). This treatment did not kill cancer cells *in vitro*, nor did it inhibit the growth of primary tumours in mice. However, when mice that had undergone tumour-removal surgery were treated with the drugs, the migration of myeloid cells to the lungs and the formation of lung tumours were reduced (Fig. 1) compared with the effects in mice that did not receive the drugs. Lu and colleagues report that a similar treatment is well tolerated by people who have lung cancer, and preliminary evidence presented by the authors also suggests that it might limit tumour recurrence after surgery.

Myeloid cells are made in the bone marrow and released into the bloodstream. Lu and colleagues report that epigenetic therapy modulated the gene expression of myeloid cells in the bone marrow of mice. By altering the chemical structure of DNA, epigenetic drugs can modify the ability of transcription-factor proteins to bind DNA and promote the expression of certain genes.

Epigenetic therapy suppressed the activity of a transcription factor called NF- κ B in the marrow-resident myeloid cells. NF- κ B controls the expression of CCR2 (which is the receptor for CCL2) and CXCR2 (which is the receptor for another tumour-derived protein called CXCL1). After treatment, monocytes expressed lower levels of CCR2, and neutrophils expressed lower levels of CXCR2. CCL2 and CXCL1 drive the recruitment and retention of myeloid cells in the lung^{4,5}. Therefore, drug-induced inhibition of CCR2 and CXCR2 expression probably impaired the process that enables a pre-metastatic niche in the lung to be populated by tumour-promoting myeloid cells.

Lu *et al.* observed that the drug-mediated epigenetic reprogramming of myeloid cells was maintained during their exit from the bone marrow and when they reached the lung. Indeed, in addition to NF- κ B, the activity of several transcription factors known to

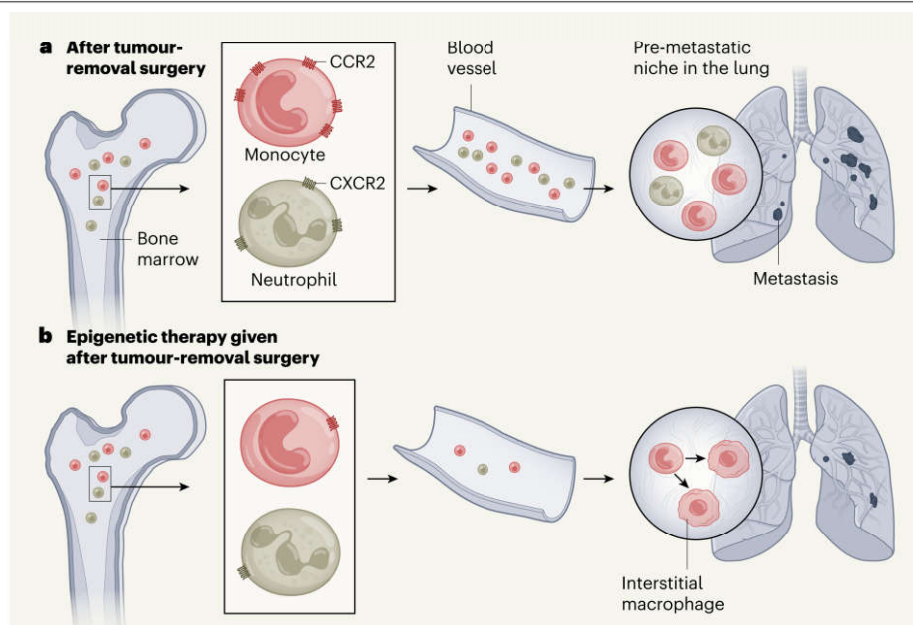


Figure 1 | Drug therapy that blocks cancer spread in mice. Even if a tumour is surgically removed, remaining cancer cells might spread to a distant organ such as the lung. **a**, Non-cancerous cells called monocytes and neutrophils, which express the proteins CCR2 and CXCR2, respectively, arise in the bone marrow. Their numbers increase in the bloodstream if the cells are mobilized by tumour-derived factors (factors not shown). The cells can travel through blood vessels to reach the lungs and form a site called a pre-metastatic niche. This site aids the establishment of cancer cells there (called a metastasis). **b**, After animals had undergone tumour-removal surgery, Lu *et al.*² treated them with epigenetic therapy, which consisted of two drugs that target the chemical structure of DNA. This treatment decreased the expression of CCR2 in monocytes and CXCR2 in neutrophils, and fewer of these cells reached the lung. The epigenetic therapy also caused some monocytes to differentiate into a cell called an interstitial macrophage. The impaired recruitment of tumour-promoting monocytes and neutrophils to the lungs was accompanied by a decrease in the number and size of metastases.

orchestrate the generation of a pre-metastatic niche⁹ was suppressed in the myeloid cells that made their way to the lung. Intriguingly, the authors found that the epigenetic therapy caused monocytes to differentiate into a type of cell called an interstitial macrophage, which might lack the ability to promote tumour establishment at a secondary site. The drugs also caused this type of cellular conversion in monocytes that were experimentally transferred from the bone marrow of untreated mice to the bloodstream of drug-treated mice. How the drugs drive this change is unknown. Perhaps their ability to rewire a cell's transcriptional network underlies this phenomenon.

Lu and colleagues' findings indicate that the re-education of tumour-hijacked myeloid cells using epigenetic therapy might be a promising strategy for thwarting their metastasis-promoting capacity. These cells have several tumour-promoting functions, such as boosting the formation of blood vessels that aid tumour growth or suppressing antitumoural T cells of the immune system^{5,9}. Thus, rather than having to individually target factors that aid different processes needed for tumour growth, epigenetic therapy might be a way to neutralize the broad and multifaceted capacity of myeloid cells to aid cancer.

Drugs such as 5-azacytidine and entinostat are being tested in the clinic for use in

combination with other anticancer therapies, both before tumour-removal surgery and in the setting of adjuvant therapy after surgery¹⁰. These drugs are not approved yet for treating solid (non-blood-cell) tumours in the clinic. High doses of such drugs are associated with severe toxicity, so using low doses that suffice to reprogram myeloid cells, together with other drugs that target cancer cells, might provide an effective anticancer treatment that avoids unwanted toxicity.

Ali Ghasemi and **Michele De Palma** are at the Swiss Institute for Experimental Cancer Research, School of Life Sciences, Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland. e-mail: michele.depalma@epfl.ch

- Swierczak, A. & Pollard, J. W. *Cold Spring Harb. Perspect. Med.* <https://doi.org/10.1101/cshperspect.a038026> (2019).
- Lu, Z. *et al.* *Nature* **579**, 284–290 (2020).
- Peinado, H. *et al.* *Nature Rev. Cancer* **17**, 302–317 (2017).
- Cassetta, L. & Pollard, J. W. *Nature Rev. Drug Discov.* **17**, 887–904 (2018).
- Gabrilovich, D. I. *Cancer Immunol. Res.* **5**, 3–8 (2017).
- Jones, P. A., Issa, J.-P. J. & Bayliss, S. *Nature Rev. Genet.* **17**, 630–641 (2016).
- Orillion, A. *et al.* *Clin. Cancer Res.* **23**, 5187–5201 (2017).
- Kim, K. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 11774–11779 (2014).
- De Palma, M., Bizziato, D. & Petrova, T. V. *Nature Rev. Cancer* **17**, 457–474 (2017).
- Cheng, Y. *et al.* *Signal Transduct. Targeted Ther.* **4**, 62 (2019).

This article was published online on 26 February 2020.

Protein synthesis

DNA-repair enzyme turns to translation

Alan J. Warren

A key DNA-repair enzyme has a surprising role during the early steps in the assembly of ribosomes – the molecular machines that translate the genetic code into protein. **See p.291**

Every minute, each human cell constructs up to 7,500 ribosomes – essential intracellular factories that decode instructions from genes to make all the proteins in the body. Ribosomes are assembled from four distinct ribosomal RNA (rRNA) molecules and 80 different proteins, which form small and large subunits, in a complex process involving more than 200 assembly factors. A better understanding of the underlying mechanisms might help to explain the devastating consequences of genetic mutations known as ribosomopathies that affect this assembly pathway. On page 291, Shao *et al.*¹ identify an unexpected role for the enzyme DNA-dependent protein kinase (DNA-PK) – a core component of the machinery for repairing DNA double-strand breaks (DSBs) – in the early steps of ribosome assembly.

Cells must repair DSBs promptly, because they threaten genomic stability and can lead to cell death or cancer. Non-homologous end joining (NHEJ) is a main pathway for DSB repair. A dimeric protein complex called KU initiates this process by binding to the broken DNA ends, then recruiting the DNA-PK catalytic subunit (DNA-PKcs) to form the active DNA-PK enzyme (Fig. 1a). DNA-PK, through its kinase activity, adds phosphate groups to the side chains of serine and threonine amino acids in other proteins, and heavily regulates itself by phosphorylating a cluster of amino acids near its serine 2056 (S2056) and threonine 2609 (T2609) residues. This activity leads to the recruitment of other enzymes, such as Artemis, that process and join the broken DNA strands.

In a comprehensive series of genetic experiments, Shao and colleagues established that both the kinase activity of DNA-PKcs and phosphorylation at its T2609 cluster are crucial for blood development (haematopoiesis) in mice. Mice that entirely lacked both DNA-PKcs and the tumour-suppressor protein p53 developed a type of blood cancer and died. By contrast, animals that did not have p53 and carried a mutant form of

DNA-PKcs lacking kinase activity survived. However, they developed a disease of the bone marrow reminiscent of a blood cancer called myelodysplastic syndrome. Moreover, mice in which amino-acid residues in the T2609 cluster were replaced by alanine residues (which could not be phosphorylated) died at four weeks old and had severe p53-dependent anaemia associated with reduced protein synthesis. This condition was reminiscent of the ribosomopathy Diamond-Blackfan anaemia (DBA), which is caused by mutations in any one of 18 different ribosomal proteins².

Shao *et al.* showed that deletion of the KU

protein completely restored haematopoiesis in mice that had mutations in the T2609 cluster, ruling out defective DNA repair alone as the explanation for the blood disorders. What, then, might DNA-PK be doing in this context?

The first precursor of the small ribosomal subunit, known as the small-subunit processome, is assembled around an RNA called U3 (ref. 3; Fig. 1b) in a subcellular compartment called the nucleolus. Shao and colleagues confirmed previous reports^{4,5} that a proportion of KU and DNA-PKcs resides in the nucleolus. These observations suggested a link between KU, DNA-PKcs and ribosome assembly. The authors provided evidence that supports this link by using U3 as ‘bait’ to identify components of the small-subunit processome, which included DNA-PKcs and KU, but not other NHEJ factors.

The small ribosomal subunit is partly comprised of an rRNA called 18S. The researchers found that unprocessed precursors of 18S rRNA accumulated in cells that lacked DNA-PKcs kinase activity, but did not accumulate when KU was deleted, too. Moreover, mice and cell lines lacking DNA-PKcs kinase activity showed reduced global protein synthesis. The authors used a technique called infrared crosslinking immunoprecipitation (irCLIP) to track down DNA-PKcs and KU to a specific location of the processome, near U3. Finally, they found that a structured fragment of U3

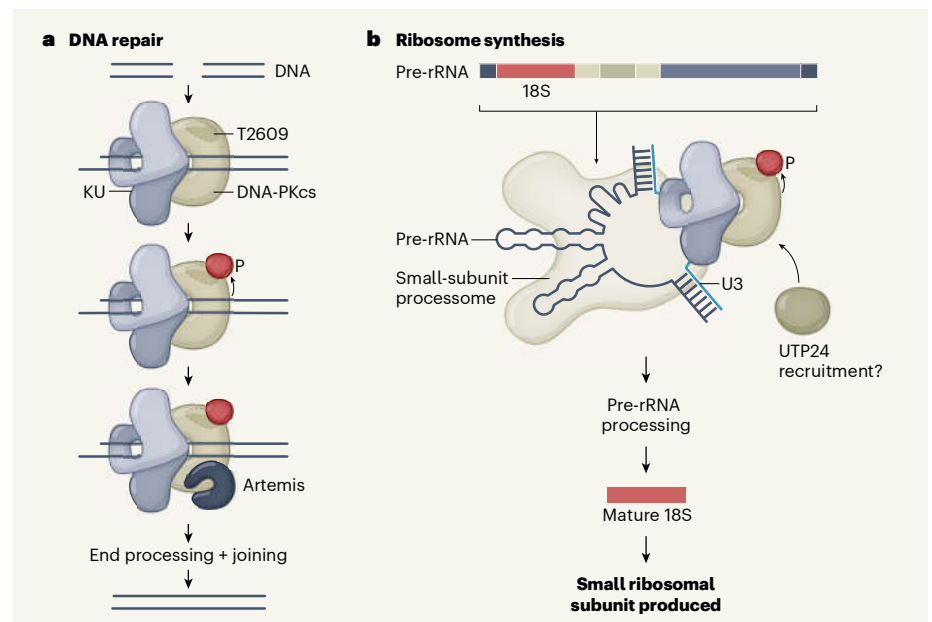


Figure 1 | Two roles for DNA-dependent protein kinase. a, To repair DNA double-strand breaks, the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) is recruited to DNA ends by the KU protein dimer. DNA-PKcs phosphorylates itself (P) on an amino-acid cluster near its threonine 2609 (T2609) residue. This enables the DNA-cleaving enzyme Artemis to access broken DNA ends, which are processed and joined. **b**, Shao *et al.*¹ have found another role for DNA-PKcs: in the synthesis of the cell's protein-producing factory, the ribosome. Precursor ribosomal RNA (pre-rRNA), which contains a region dubbed 18S, forms part of the ribosomal small-subunit processome. The authors find that KU recruits DNA-PKcs to another RNA molecule in the processome, U3. Self-phosphorylation might trigger an RNA-dependent conformational change in DNA-PKcs, regulating access of an RNA-cleaving enzyme such as UTP24, which cleaves the pre-rRNA to produce mature 18S rRNA that forms part of the ribosome.

drives the assembly of DNA-PK and stimulates its catalytic activity *in vitro*, although does so much less efficiently than can DNA.

Taken together, these observations suggest a model in which KU recruits DNA-PKs to the small-subunit processome. In the case of kinase-defective DNA-PK, the mutant enzyme's inability to regulate its own activity gives the protein a new function, blocking the processing of precursor rRNA into mature 18S rRNA in the small-subunit processome. The resulting defect in global protein synthesis drives a p53-dependent loss of red-blood-cell precursors – a cell type that has an especially high physiological demand for protein synthesis. The parallels with NHEJ are intriguing: in that pathway, the complete deletion of DNA-PKs results in only a minor reduction in repair fidelity, and the joining of broken DNA ends is retained. By contrast, the kinase-inactive DNA-PKs mutant is wholly unable to carry out end joining.

The specific role of DNA-PK in precursor rRNA processing, and how it recognizes precursor rRNA *in vivo*, remains unclear. However, structural analysis of the yeast small-subunit processome⁶ has revealed that U3 acts as a molecular guide that docks the processome onto the precursor rRNA by forming four evolutionarily conserved duplexes (hinges) between the two components: two hinges in a highly branched region of the precursor rRNA, and two in a region that will become the mature 18S rRNA. These hinges are a prerequisite for three cleavage events, mediated by an RNA-cleaving nuclease enzyme, that release the 18S rRNA ready to make the small subunit.

Shao *et al.* show that DNA-PK and KU primarily interact with U3 at this hinge region. Thus, much as DNA-PKs recruits the DNA-cleaving enzyme Artemis during the NHEJ processing of DNA ends⁷, with U3, DNA-PKs might also help to recruit specific RNA-cleaving nucleases (such as UTP24) to the small-subunit processome to cleave the precursor rRNA for ribosome construction.

Structural studies suggest that the binding of DNA-PKs to KU and DNA could regulate the activation of DNA-PKs kinase activity allosterically, that is, by changing the conformation of the enzyme^{8–10}. In the future, it will be interesting to compare RNA- and DNA-dependent conformational changes in DNA-PKs. The physiological relevance of the broad array of RNA partners identified by Shao *et al.* in their iCLIP analysis also remains to be dissected.

Shao and colleagues' study has identified an interesting player in ribosome assembly that might efficiently couple DNA DSB repair with processing of precursor rRNA, which is highly transcribed from the naturally unstable ribosomal DNA template. Broadly, the findings encourage us to critically evaluate how dynamic redistribution of DNA-PK might allow the cell to couple DSB repair with the

regulation of protein synthesis. And, although further studies are required, we might have taken a step closer to deciphering the mysterious ribosomopathies.

Alan J. Warren is at the Cambridge Institute for Medical Research, Hills Road, Cambridge CB2 0XY, UK.
e-mail: ajw1000@cam.ac.uk

1. Shao, Z. *et al.* *Nature* **579**, 291–296 (2020).
2. Khajuria, R. K. *et al.* *Cell* **173**, 90–103 (2018).

3. Dragon, F. *et al.* *Nature* **417**, 967–970 (2002).
4. Adelmant, G. *et al.* *Mol. Cell. Proteom.* **11**, 411–421 (2012).
5. Britton, S., Coates, J. & Jackson, S. P. *J. Cell Biol.* **202**, 579–595 (2013).
6. Barandun, J. *et al.* *Nature Struct. Mol. Biol.* **24**, 944–953 (2017).
7. Ma, Y. *et al.* *Cell* **108**, 781–794 (2002).
8. Yin, X. *et al.* *Cell Res.* **27**, 1341–1350 (2017).
9. Sharif, H. *et al.* *Proc. Natl Acad. Sci. USA* **114**, 7367–7372 (2017).
10. Sibanda, B. L. *et al.* *Science* **355**, 520–524 (2017).

This article was published online on 26 February 2020.

Palaeontology

Tiny fossil sheds light on miniaturization of birds

Roger B. J. Benson

A tiny skull trapped in 99-million-year-old amber suggests that some of the earliest birds evolved to become miniature. The fossil illustrates how ancient amber can act as a window into the distant past. **See p.245**

Dinosaurs were big, whereas birds – which evolved from dinosaurs – are small. This variation is of great importance, because body size affects lifespan, food requirements, sensory capabilities and many other fundamental aspects of biology. The smallest dinosaurs¹ weighed hundreds of grams, but the smallest living bird, the bee hummingbird (*Mellisuga helenae*)², weighs only 2 grams. How did this difference come about, and why? On page 245, Xing *et al.*³ describe the tiny, fossilized, bird-like skull of a previously unknown species, which they name *Oculudentavis khaungrae*.

The discovery suggests that miniature body sizes in birds evolved earlier than previously recognized, and might provide insights into the evolutionary process of miniaturization.

Fossilization of bones in sediments such as clay, silt and sand can crush and destroy the remains of small animals, and can flatten and decay soft parts such as skin, scales and feathers. By contrast, preservation of small animals in Burmese amber (which formed from the resin flows of coniferous trees about 99 million years ago) helps to protect their soft parts. A wide range of invertebrates⁴ and small



Figure 1 | Computed tomography scan of the skull of *Oculudentavis khaungrae*. Xing *et al.*³ have characterized this 99-million-year-old fossil bird.

drives the assembly of DNA-PK and stimulates its catalytic activity *in vitro*, although does so much less efficiently than can DNA.

Taken together, these observations suggest a model in which KU recruits DNA-PKs to the small-subunit processome. In the case of kinase-defective DNA-PK, the mutant enzyme's inability to regulate its own activity gives the protein a new function, blocking the processing of precursor rRNA into mature 18S rRNA in the small-subunit processome. The resulting defect in global protein synthesis drives a p53-dependent loss of red-blood-cell precursors – a cell type that has an especially high physiological demand for protein synthesis. The parallels with NHEJ are intriguing: in that pathway, the complete deletion of DNA-PKs results in only a minor reduction in repair fidelity, and the joining of broken DNA ends is retained. By contrast, the kinase-inactive DNA-PKs mutant is wholly unable to carry out end joining.

The specific role of DNA-PK in precursor rRNA processing, and how it recognizes precursor rRNA *in vivo*, remains unclear. However, structural analysis of the yeast small-subunit processome⁶ has revealed that U3 acts as a molecular guide that docks the processome onto the precursor rRNA by forming four evolutionarily conserved duplexes (hinges) between the two components: two hinges in a highly branched region of the precursor rRNA, and two in a region that will become the mature 18S rRNA. These hinges are a prerequisite for three cleavage events, mediated by an RNA-cleaving nuclease enzyme, that release the 18S rRNA ready to make the small subunit.

Shao *et al.* show that DNA-PK and KU primarily interact with U3 at this hinge region. Thus, much as DNA-PKs recruits the DNA-cleaving enzyme Artemis during the NHEJ processing of DNA ends⁷, with U3, DNA-PKs might also help to recruit specific RNA-cleaving nucleases (such as UTP24) to the small-subunit processome to cleave the precursor rRNA for ribosome construction.

Structural studies suggest that the binding of DNA-PKs to KU and DNA could regulate the activation of DNA-PKs kinase activity allosterically, that is, by changing the conformation of the enzyme^{8–10}. In the future, it will be interesting to compare RNA- and DNA-dependent conformational changes in DNA-PKs. The physiological relevance of the broad array of RNA partners identified by Shao *et al.* in their iCLIP analysis also remains to be dissected.

Shao and colleagues' study has identified an interesting player in ribosome assembly that might efficiently couple DNA DSB repair with processing of precursor rRNA, which is highly transcribed from the naturally unstable ribosomal DNA template. Broadly, the findings encourage us to critically evaluate how dynamic redistribution of DNA-PK might allow the cell to couple DSB repair with the

regulation of protein synthesis. And, although further studies are required, we might have taken a step closer to deciphering the mysterious ribosomopathies.

Alan J. Warren is at the Cambridge Institute for Medical Research, Hills Road, Cambridge CB2 0XY, UK.
e-mail: ajw1000@cam.ac.uk

1. Shao, Z. *et al.* *Nature* **579**, 291–296 (2020).
2. Khajuria, R. K. *et al.* *Cell* **173**, 90–103 (2018).

3. Dragon, F. *et al.* *Nature* **417**, 967–970 (2002).
4. Adelmant, G. *et al.* *Mol. Cell. Proteom.* **11**, 411–421 (2012).
5. Britton, S., Coates, J. & Jackson, S. P. *J. Cell Biol.* **202**, 579–595 (2013).
6. Barandun, J. *et al.* *Nature Struct. Mol. Biol.* **24**, 944–953 (2017).
7. Ma, Y. *et al.* *Cell* **108**, 781–794 (2002).
8. Yin, X. *et al.* *Cell Res.* **27**, 1341–1350 (2017).
9. Sharif, H. *et al.* *Proc. Natl Acad. Sci. USA* **114**, 7367–7372 (2017).
10. Sibanda, B. L. *et al.* *Science* **355**, 520–524 (2017).

This article was published online on 26 February 2020.

Palaeontology

Tiny fossil sheds light on miniaturization of birds

Roger B. J. Benson

A tiny skull trapped in 99-million-year-old amber suggests that some of the earliest birds evolved to become miniature. The fossil illustrates how ancient amber can act as a window into the distant past. **See p.245**

Dinosaurs were big, whereas birds – which evolved from dinosaurs – are small. This variation is of great importance, because body size affects lifespan, food requirements, sensory capabilities and many other fundamental aspects of biology. The smallest dinosaurs¹ weighed hundreds of grams, but the smallest living bird, the bee hummingbird (*Mellisuga helenae*)², weighs only 2 grams. How did this difference come about, and why? On page 245, Xing *et al.*³ describe the tiny, fossilized, bird-like skull of a previously unknown species, which they name *Oculudentavis khaungrae*.

The discovery suggests that miniature body sizes in birds evolved earlier than previously recognized, and might provide insights into the evolutionary process of miniaturization.

Fossilization of bones in sediments such as clay, silt and sand can crush and destroy the remains of small animals, and can flatten and decay soft parts such as skin, scales and feathers. By contrast, preservation of small animals in Burmese amber (which formed from the resin flows of coniferous trees about 99 million years ago) helps to protect their soft parts. A wide range of invertebrates⁴ and small



Figure 1 | Computed tomography scan of the skull of *Oculudentavis khaungrae*. Xing *et al.*³ have characterized this 99-million-year-old fossil bird.

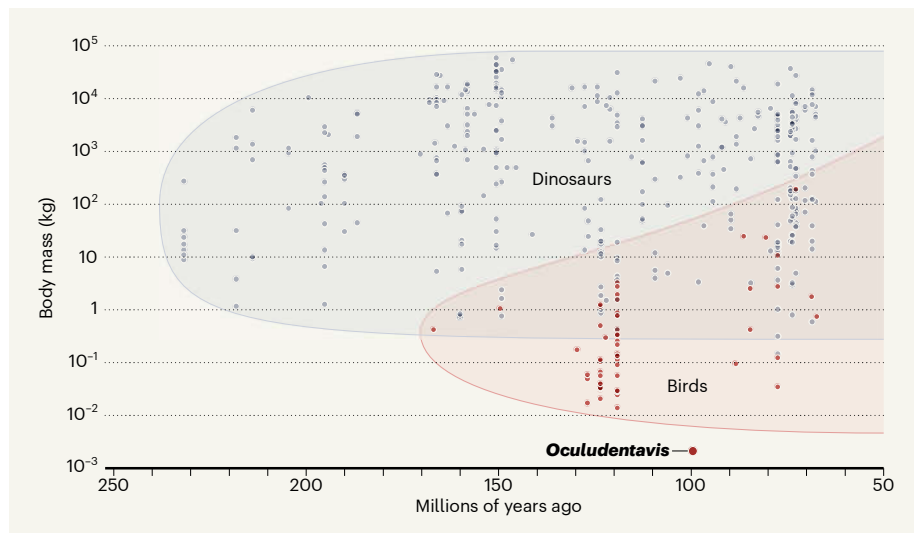


Figure 2 | Different size ranges of dinosaurs and birds. Dinosaurs varied from about 500 grams to many tonnes in weight. By contrast, the first birds were much smaller. The smallest fossil bird found so far from the Cretaceous period weighs in at about 12 grams (data taken from ref. 9). Xing *et al.*³ report that the tiny *Oculudentavis* weighed just 2 grams. This discovery provides new insight into the lower limits of vertebrate body size in the age of dinosaurs.

vertebrates, including lizards⁵ and birds⁶, have been found in Burmese amber. Specimens preserved in this material are rapidly emerging as an exceptional way to study tiny vertebrates from the age of dinosaurs^{5,6}.

It is in Burmese amber that the single known fossil skull of *Oculudentavis* has been preserved (see Fig. 1a of the paper³). *Oculudentavis* means eye tooth bird, a name that Xing *et al.* chose because of two unusual features of the skull, each of which provides evidence about the likely lifestyle of this 99-million-year-old species.

First, the skull is dominated by two enormous eye sockets containing scleral ossicles – rings of bone that form the eye skeletons of birds (Fig. 1). The opening at the centre of these ossicles is narrow, restricting access for light into the eye and providing strong evidence that *Oculudentavis* was active in well-lit, daytime environments.

Second, the jaws of *Oculudentavis* have many small teeth. This might seem odd, given the absence of teeth in today's birds, but teeth are in fact common among early fossil birds⁷. However, *Oculudentavis* has more teeth than other birds of the period, and these extend unusually far back in the jaws to a point just under the eye. On the basis of these facts, along with observations of the fossilized tongue, the authors suggest that *Oculudentavis* was a predator that mainly ate invertebrates. This diet differs considerably from the nectar-based diet of the smallest living birds, and suggests that extinct and living birds took different paths to miniaturization (although how diet might be involved in this process remains unknown).

Oculudentavis is just one fossil species. However, even single fossils can contribute

greatly to our understanding of the history of life on Earth. In this case, weighing perhaps 2 grams, *Oculudentavis* is about one-sixth of the size of the smallest known early fossil bird¹. This indicates that, only shortly after their origins late in the Jurassic period (which lasted from about 201 million to 145 million years ago), birds had already attained their minimum body sizes. By contrast, the smallest dinosaurs weighed hundreds of times more¹ (Fig. 2). Understanding when, how and why the lower limits of body size shifted in this way requires greater knowledge of the earliest fossil birds. But *Oculudentavis* is a stepping stone towards this.

The evolutionary relationships between

“Even single fossils can contribute greatly to our understanding of the history of life on Earth.”

Oculudentavis and other dinosaurs and birds are difficult to determine, but are central to clarifying the evolutionary implications of this discovery. Xing and colleagues' analysis suggests two possibilities. *Oculudentavis* could belong to the most common group of birds of the Cretaceous period (about 145 million to 66 million years ago), the enantiornithines. Alternatively, it could be much more closely related to dinosaurs, lying almost midway on the evolutionary tree between the Cretaceous birds and *Archaeopteryx*, the iconic winged dinosaur from the Jurassic.

This confusion is a result of the bizarre features seen in *Oculudentavis*. These include many characteristics that differ from those of

other birds, such as more-robust, fused bones, and proportionally enlarged sensory organs relative to the overall body size. The authors suggest that these features could have arisen from the constraints of evolutionary miniaturization or from ecological specialization. Both of these might have required *Oculudentavis* to have a strengthened skull and proportionally large eyes to maintain sensory capacity at such a tiny size. In addition, *Oculudentavis* has features that are not seen in dinosaurs or birds, but are present in lizards – these include the spoon shape of its scleral ossicles and the fact that its teeth are attached to the jaw bone by their sides, rather than being implanted in sockets. The challenge of determining how *Oculudentavis* is related to other early birds and bird-like dinosaurs would be greatly assisted by knowing more about its skeleton.

The past decade has generated much data on the dinosaur–bird transition, greatly advancing our understanding of this major evolutionary event^{7,8}. In the past few years, Burmese amber has yielded surprising insights, including previously unseen feather and skeletal structures in other extinct birds⁶. The study of small vertebrates preserved in amber, their ecosystems and their evolutionary relationships with one another is in a nascent phase. But *Oculudentavis* suggests that the potential for continued discovery remains large – especially for animals of diminutive sizes.

Roger B. J. Benson is in the Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK.
e-mail: roger.benson@earth.ox.ac.uk

1. Benson, R. B. J., Hunt, G., Carrano, M. T. & Campione, N. *Palaeontol.* **61**, 13–48 (2018).
2. Del Hoyo, J., Elliott, A. & Sargatal, J. (eds) *Handbook of the Birds of the World Vol. 5* (Lynx, 1999).
3. Xing, L. *et al.* *Nature* **579**, 245–249 (2020).
4. Grimaldi, D. A., Engel, M. S. & Nascimbene, P. C. *Am. Mus. Novit.* **3361**, 1–71 (2002).
5. Daza, J., Stanley, E. L., Wagner, P., Bauer, A. M. & Grimaldi, D. A. *Sci. Adv.* **2**, e1501080 (2016).
6. Xing, L., McKellar, R. C., O'Connor, J. K., Nou, K. & Mai, H. *Sci. Rep.* **9**, 15513 (2019).
7. O'Connor, J. K., Chiappe, L. M. & Bell, A. in *Living Dinosaurs: The Evolutionary History of Modern Birds* (eds Dyke, G. & Kaiser, G.) Ch. 3, 39–114 (Wiley-Blackwell, 2011).
8. Xu, X. *et al.* *Science* **346**, 1253293 (2014).
9. Benson, R. B. J. *et al.* *PLoS Biol.* **12**, e1001853 (2014).

An electrically operated magnetic logic gate

See-Hun Yang

Bits of a logic gate can be encoded by differently magnetized regions. A method has been developed in which the walls between these domains are manipulated electrically, rather than magnetically, to produce a logic gate. **See p.214**

Computer logic based on standard electronic components is expected to hit its speed and power limits before long. To get around this, logic devices that use magnetic elements, rather than electronic ones, have been proposed – but these usually require external magnetic fields, which limits their application. A working logic gate that uses magnetic elements and is driven completely electrically, without the need for an external magnetic field, has yet to be demonstrated¹. However, on page 214, Luo *et al.*² report that they have harnessed the chirality ('handedness') of a system to invert the direction of magnetization of domains in a cobalt wire purely by means of an electric current. The resulting inverter device acts as a 'NOT' gate, which they use to build up other logic gates such as NAND and NOR.

First, some background. Magnetic domains are small regions of uniform magnetization in a material. The narrow boundaries separating different magnetization orientations are called domain walls; within the wall, the magnetization must gradually twist around. Such magnetic domains have long been used to encode data bits for storage, such as in hard disk drives, in which mechanical motion is needed to access the data.

By contrast, in magnetic racetrack memory³ (which is still under development), no mechanical movement is needed. Instead, magnetic domains within hair-thin metal wires are moved by the flow of electric current in the wires. The trick to doing this lies in an interaction – the spin-transfer torque – between the magnetic moments of the domain walls and the spins of the moving electrons. Here, spin refers to a quantum property of the electrons; moment is the magnetic direction and strength; and the torque is a twisting interaction that tends to rotate the moments and thus move the walls.

In spin-transfer systems, the torque arises from current in the magnetic wire itself. But another type of torque, known as spin-orbit torque, is produced when the wire is placed on a layer of a non-magnetic heavy metal such

as platinum, and the electric current flows in that layer instead (Fig. 1). In such systems, two effects add to each other to drive twisting in the domain walls more efficiently than in spin-transfer systems^{4,5}. The first is that a spin current arises in the non-magnetic heavy-metal layer; the second effect, which is crucial to Luo and colleagues' work, is that an interaction between the two metals forces a specific chirality on the domain walls.

Chirality means handedness: like left and right hands, objects of opposite chirality have mirror symmetry, but cannot be superimposed on each other. Domain walls known as Néel-type walls can exhibit chirality when the direction of their magnetic moment is reversed, because there are two ways in which reversal can happen: the moment can undergo a right-handed twist or a left-handed one. (Moments can also spin in the plane of domain walls known as Bloch walls, but this mechanism is not used in Luo and co-workers' study.)

The chirality is produced by an effect called the Dzyaloshinskii–Moriya exchange interaction (DMI), which acts between the magnetic and non-magnetic metals. The DMI both establishes a specific chirality and favours the generation of the correct type of wall – a Néel wall – in Luo and colleagues' set-up. Researchers have previously recognized that chiral domain walls could be used in logic operations⁶, but an all-electrical working logic has been hampered by the lack of a reliable working inverter, the key component for logic operations.

Luo *et al.* have ingeniously invented an inverter that flips the magnetic moments associated with incoming bits of data using spin-orbit torque. Building on their previous

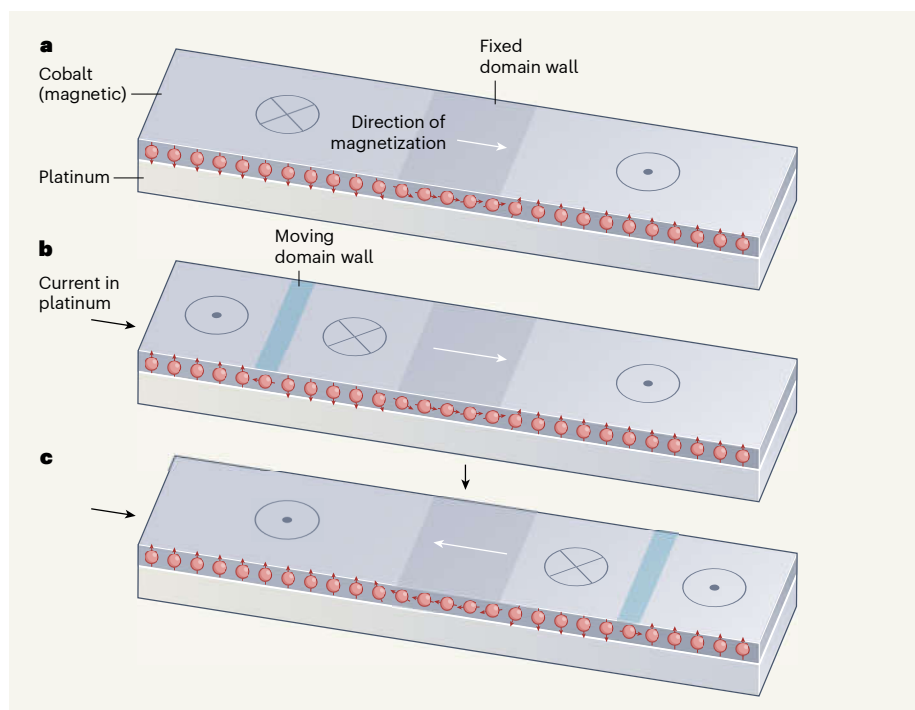


Figure 1 | How to flip magnetic data bits electrically. Luo and co-workers² have produced a NOT logic gate that flips the direction of magnetization of magnetic domains by electrical means. In the authors' system, mobile domains of a cobalt wire that have up or down magnetic moments act as data bits. From the side, spheres with arrows represent the direction of magnetization; on the top surface, circled dots and crosses indicate up and down moments, respectively. **a**, The authors fabricated a fixed wall between up and down domains, in which the magnetization direction can rotate in a left- or right-handed direction – a property called chirality. **b**, An electric current in the platinum substrate propels another, moving chiral domain wall (blue) along the wire. **c**, When two opposite magnetic moments collide on the left-hand side of the fixed wall, the direction of the moment in the fixed region switches. A new domain forms on the other side of the wall, in which the moment is reversed, preserving the preferred chirality of the system. Overall, the bit has been flipped. (Adapted from Fig. 1 of the paper².)

work on chirally coupled nanomagnets⁷, the authors fabricated a sort of artificial, stationary domain wall in a magnetic cobalt wire interfaced with non-magnetic platinum. The magnetization in the cobalt is perpendicular to the plane of the wire, except in the stationary region. There, it is magnetized in the direction of the wire's long axis, like the region in the middle of an ordinary domain wall, but across a much larger width. This is crucial, because it allows smaller coercivity – that is, the magnetization here is easier to switch.

To picture how the inverter works, consider an input consisting of a domain wall that has left-handed chirality (Fig. 1). This mobile wall is rolled along the wire by spin-orbit torque. When it reaches the fixed artificial boundary, two opposite magnetic moments collide, producing a region of the wire in which the moment changes abruptly. According to theories of magnetism, such an abrupt change has a high energy cost. To lower the energy of the system, one of the moments must be switched, or a new magnetic domain must be generated. In this case, the moment in the low-coercivity fixed wall switches to the same direction as that in the incoming wall.

But a chirality effect now comes into play: this switch of magnetic moment produces a right-handed chirality at the other side of the fixed wall that conflicts with the chirality preferred by the DMI. To resolve this, a new domain wall forms on that side (the system is shaped in such a way as to promote this process) and sets off along the wire. The moments in the resulting outgoing bit thus have the preferred left-handed chirality, rather than the right-handed chirality originally produced at the wall.

By integrating their inverters into junctions, Luo *et al.* designed some simple logic gates (NAND and NOR), as well as more-complicated ones (such as exclusive-OR). Each junction has two inputs, an intrinsic bias towards one magnetic moment and one output. The output is determined by the two inverted inputs and by the bias at the junctions (rather like a 'majority gate'⁸). So, when the inputs and bias are (0,0) and 1, respectively, inverters immediately before the junction invert them to (1,1) and 0 at the junction itself, which consequently outputs 1, acting as a NOT gate. But when the inputs are either (1,0) or (0,1), the value of the bias determines whether the gate behaves as a NOR or a NAND. This majority-gate behaviour mitigates the need to precisely synchronize the two inputs, offering reliable logic operations.

This logic system satisfies key criteria known as cascability and fan-out. Cascability means that the output of one gate is produced in the correct form and is strong enough to drive the input of the next gate. And fan-out means that one gate output can be connected to several gate inputs⁹. Moreover, the

data can be stored in the absence of an external power source, and evade damage by ionizing radiation.

Challenges remain before chips based on Luo and co-workers' system can reach the market. The operating current will need to be reduced so that it can be accommodated by tiny complementary metal oxide semiconductor (CMOS) transistors, which help to pick up inputs and outputs for use in chips. In theory, current decreases as the size of wires and transistors decreases, and so current density (the charge per unit time that flows through a given cross-section of the wire) remains constant with scaling. A reduction in current density will be needed to increase the speed and reduce the energy consumption of the authors' system. Domain-wall velocity does not scale linearly with current, and so new materials might need to be used to reduce the current density^{10,11}.

Another issue is that the input and output states of Luo and colleagues' system have to be detected by microscopy, rather than by an electrical method. A different read-out system

will be needed for practical applications, but this could be technically challenging. An effect known as tunnelling magneto-resistance might offer one solution¹². The implementation of a domain-wall logic chip that uses an electrically driven read-out system should be the next goal, following on from Luo and colleagues' exciting discovery.

See-Hun Yang is at IBM Research – Almaden, San Jose, California 95120, USA.
e-mail: seeyang@us.ibm.com

1. Allwood, D. A. *et al.* *Science* **309**, 1688–1692 (2005).
2. Luo, Z. *et al.* *Nature* **579**, 214–218 (2020).
3. Parkin, S. & Yang, S.-H. *Nature Nanotechnol.* **10**, 195–198 (2015).
4. Ryu, K.-S., Thomas, L., Yang, S.-H. & Parkin, S. *Nature Nanotechnol.* **8**, 527–533 (2013).
5. Emori, S., Bauer, U., Ahn, S.-M., Martinez, E. & Beach, G. S. D. *Nature Mater.* **12**, 611–616 (2013).
6. Omari, K. A. *et al.* *Adv. Funct. Mater.* **29**, 1807282 (2019).
7. Luo, Z. *et al.* *Science* **363**, 1435–1439 (2019).
8. Imre, A. *et al.* *Science* **311**, 205–208 (2006).
9. Miller, D. A. B. *Nature Photon.* **4**, 3–5 (2010).
10. Yang, S.-H., Ryu, K.-S. & Parkin, S. *Nature Nanotechnol.* **10**, 221–226 (2015).
11. Avci, C. O. *et al.* *Nature Nanotechnol.* **14**, 561–566 (2019).
12. Julliere, M. *Phys. Lett. A* **54**, 225–226 (1975).

Neuroscience

Research on the cerebellum yields rewards

Jennifer L. Raymond

A brain structure called the cerebellum has mostly been associated with learning from errors. The discovery that the cerebellum is also involved in reward-driven learning in monkeys implies a previously unappreciated role in cognition.

People and organizations alike use rewards, from snack to salary bonuses and frequent-flyer miles, to shape behaviour through a process called reinforcement learning. For example, if a dog receives a treat for rolling over in response to a verbal command, the likelihood of that behavioural response to the verbal cue will increase. Writing in *Neuron*, Sendhilnathan and colleagues¹ describe neuronal signals that could support such reward-driven learning. What is remarkable is where the authors found these signals – not in the brain areas that have long been implicated in reinforcement learning, but in the cerebellum, a brain structure historically associated with error-driven, rather than reward-driven, learning.

The cerebellum is best known for its role in motor-skill learning – the process by which movements become smooth and accurate through practice. Fifty years of research² supports the idea that when you practise a movement, such as your tennis backhand,

the cerebellum uses feedback about errors to gradually refine the accuracy of the movement by weakening the neuronal connections that are responsible for those errors. It has been widely assumed that the cerebellum uses a similar, error-correcting learning algorithm to support cognition³, because the regions of the cerebellum that contribute to cognitive functions such as navigation⁴ and social behaviour⁵ have the same basic circuit architecture as those that control movement.

In the past three years, however, there has been a flurry of studies showing reward-related neuronal activity in the cerebellum^{6–12}. What are reward signals doing in an error-correcting part of the brain? Sendhilnathan *et al.* leveraged the rapid learning abilities of monkeys to gain fresh insights into reward-related signalling in the cerebellum.

In each experimental session, the authors presented a monkey with two visual cues it had never seen before on a computer screen.

work on chirally coupled nanomagnets⁷, the authors fabricated a sort of artificial, stationary domain wall in a magnetic cobalt wire interfaced with non-magnetic platinum. The magnetization in the cobalt is perpendicular to the plane of the wire, except in the stationary region. There, it is magnetized in the direction of the wire's long axis, like the region in the middle of an ordinary domain wall, but across a much larger width. This is crucial, because it allows smaller coercivity – that is, the magnetization here is easier to switch.

To picture how the inverter works, consider an input consisting of a domain wall that has left-handed chirality (Fig. 1). This mobile wall is rolled along the wire by spin-orbit torque. When it reaches the fixed artificial boundary, two opposite magnetic moments collide, producing a region of the wire in which the moment changes abruptly. According to theories of magnetism, such an abrupt change has a high energy cost. To lower the energy of the system, one of the moments must be switched, or a new magnetic domain must be generated. In this case, the moment in the low-coercivity fixed wall switches to the same direction as that in the incoming wall.

But a chirality effect now comes into play: this switch of magnetic moment produces a right-handed chirality at the other side of the fixed wall that conflicts with the chirality preferred by the DMI. To resolve this, a new domain wall forms on that side (the system is shaped in such a way as to promote this process) and sets off along the wire. The moments in the resulting outgoing bit thus have the preferred left-handed chirality, rather than the right-handed chirality originally produced at the wall.

By integrating their inverters into junctions, Luo *et al.* designed some simple logic gates (NAND and NOR), as well as more-complicated ones (such as exclusive-OR). Each junction has two inputs, an intrinsic bias towards one magnetic moment and one output. The output is determined by the two inverted inputs and by the bias at the junctions (rather like a 'majority gate'⁸). So, when the inputs and bias are (0,0) and 1, respectively, inverters immediately before the junction invert them to (1,1) and 0 at the junction itself, which consequently outputs 1, acting as a NOT gate. But when the inputs are either (1,0) or (0,1), the value of the bias determines whether the gate behaves as a NOR or a NAND. This majority-gate behaviour mitigates the need to precisely synchronize the two inputs, offering reliable logic operations.

This logic system satisfies key criteria known as cascability and fan-out. Cascability means that the output of one gate is produced in the correct form and is strong enough to drive the input of the next gate. And fan-out means that one gate output can be connected to several gate inputs⁹. Moreover, the

data can be stored in the absence of an external power source, and evade damage by ionizing radiation.

Challenges remain before chips based on Luo and co-workers' system can reach the market. The operating current will need to be reduced so that it can be accommodated by tiny complementary metal oxide semiconductor (CMOS) transistors, which help to pick up inputs and outputs for use in chips. In theory, current decreases as the size of wires and transistors decreases, and so current density (the charge per unit time that flows through a given cross-section of the wire) remains constant with scaling. A reduction in current density will be needed to increase the speed and reduce the energy consumption of the authors' system. Domain-wall velocity does not scale linearly with current, and so new materials might need to be used to reduce the current density^{10,11}.

Another issue is that the input and output states of Luo and colleagues' system have to be detected by microscopy, rather than by an electrical method. A different read-out system

will be needed for practical applications, but this could be technically challenging. An effect known as tunnelling magneto-resistance might offer one solution¹². The implementation of a domain-wall logic chip that uses an electrically driven read-out system should be the next goal, following on from Luo and colleagues' exciting discovery.

See-Hun Yang is at IBM Research – Almaden, San Jose, California 95120, USA.
e-mail: seeyang@us.ibm.com

1. Allwood, D. A. *et al.* *Science* **309**, 1688–1692 (2005).
2. Luo, Z. *et al.* *Nature* **579**, 214–218 (2020).
3. Parkin, S. & Yang, S.-H. *Nature Nanotechnol.* **10**, 195–198 (2015).
4. Ryu, K.-S., Thomas, L., Yang, S.-H. & Parkin, S. *Nature Nanotechnol.* **8**, 527–533 (2013).
5. Emori, S., Bauer, U., Ahn, S.-M., Martinez, E. & Beach, G. S. D. *Nature Mater.* **12**, 611–616 (2013).
6. Omari, K. A. *et al.* *Adv. Funct. Mater.* **29**, 1807282 (2019).
7. Luo, Z. *et al.* *Science* **363**, 1435–1439 (2019).
8. Imre, A. *et al.* *Science* **311**, 205–208 (2006).
9. Miller, D. A. B. *Nature Photon.* **4**, 3–5 (2010).
10. Yang, S.-H., Ryu, K.-S. & Parkin, S. *Nature Nanotechnol.* **10**, 221–226 (2015).
11. Avci, C. O. *et al.* *Nature Nanotechnol.* **14**, 561–566 (2019).
12. Julliere, M. *Phys. Lett. A* **54**, 225–226 (1975).

Neuroscience

Research on the cerebellum yields rewards

Jennifer L. Raymond

A brain structure called the cerebellum has mostly been associated with learning from errors. The discovery that the cerebellum is also involved in reward-driven learning in monkeys implies a previously unappreciated role in cognition.

People and organizations alike use rewards, from snack to salary bonuses and frequent-flyer miles, to shape behaviour through a process called reinforcement learning. For example, if a dog receives a treat for rolling over in response to a verbal command, the likelihood of that behavioural response to the verbal cue will increase. Writing in *Neuron*, Sendhilnathan and colleagues¹ describe neuronal signals that could support such reward-driven learning. What is remarkable is where the authors found these signals – not in the brain areas that have long been implicated in reinforcement learning, but in the cerebellum, a brain structure historically associated with error-driven, rather than reward-driven, learning.

The cerebellum is best known for its role in motor-skill learning – the process by which movements become smooth and accurate through practice. Fifty years of research² supports the idea that when you practise a movement, such as your tennis backhand,

the cerebellum uses feedback about errors to gradually refine the accuracy of the movement by weakening the neuronal connections that are responsible for those errors. It has been widely assumed that the cerebellum uses a similar, error-correcting learning algorithm to support cognition³, because the regions of the cerebellum that contribute to cognitive functions such as navigation⁴ and social behaviour⁵ have the same basic circuit architecture as those that control movement.

In the past three years, however, there has been a flurry of studies showing reward-related neuronal activity in the cerebellum^{6–12}. What are reward signals doing in an error-correcting part of the brain? Sendhilnathan *et al.* leveraged the rapid learning abilities of monkeys to gain fresh insights into reward-related signalling in the cerebellum.

In each experimental session, the authors presented a monkey with two visual cues it had never seen before on a computer screen.

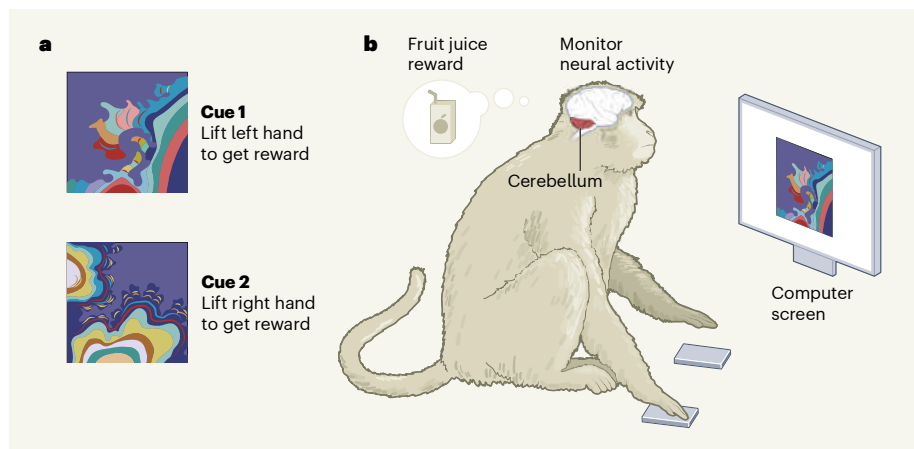


Figure 1 | A rewarding choice. **a**, Sendhilnathan *et al.*¹ examined neural activity in a brain region called the cerebellum during reward-driven learning. The authors presented monkeys with two visual cues. For one cue, the monkey needed to lift its left hand to receive a reward of fruit juice; for the other, lifting the right hand would lead to a reward. **b**, The monkeys performed a series of trials in which they were presented with one of the two cues. The authors monitored cerebellar neuronal activity while the monkeys learnt, through trial and error, which response would produce a reward for each cue. They found that a subpopulation of neurons carried information about the success or failure of the previous trial until the next trial was completed (not shown). (Figure adapted from Fig. S2 of ref. 1.)

One arbitrarily assigned cue would result in the monkey receiving a reward of fruit juice if the animal responded by lifting its left hand. The other cue would result in a reward if the monkey lifted its right hand. The researchers monitored the activity of neurons called Purkinje cells in the cerebellum as the monkeys learnt, through trial and error, to make the correct response to each visual cue (Fig. 1).

Sendhilnathan *et al.* found that the activity of cerebellar Purkinje cells carried information about the success or failure of the monkey's most recent attempt at the task. One subpopulation showed high activity following a correct response to the cue; another showed high activity following a failed attempt. These signals arose a few hundred milliseconds after the end of a trial and persisted until the next trial was completed. As such, they seemed to provide a working memory that could enable the outcome of one trial to guide the next behavioural choice.

These signals are reminiscent of those carried by neurons in frontal and parietal regions of the brain's cerebral cortex, which encode the 'value' of different behavioural choices on the basis of reward history over multiple trials¹³. In the current study, the cerebellar neurons kept track of only the most recent trial's outcome. But in this task, the outcome of a single trial provides sufficient information for the monkey to infer the correct response for the next trial – if a reward was not given when a monkey lifted its right hand in response to one visual cue, for instance, then the correct response to that cue must be to lift the left hand, and the correct response to the other visual cue would be to lift the right hand. It would be interesting to know whether cerebellar neurons can keep

track of a more-extended history of rewards should the task require it, and whether the cerebellum interacts with the cerebral cortex in performing this computation.

Importantly, information about the previous trial's outcome was present in the cerebellum only when a new set of cue–response associations was being learnt. As monkeys improved their performance over trials, the neuronal activity encoding each outcome waned. Moreover, the signal was not present when monkeys earned rewards by responding to a pair of visual cues that they had mastered through several months of training. These observations indicate that cerebellar neurons are not simply carrying information about rewards,

“The cerebellum seems to contribute specifically to learning about how to earn rewards in a new situation.”

predictions about rewards or the movements that animals make when anticipating rewards. Rather, the cerebellum seems to contribute specifically to learning about how to earn rewards in a new situation. The authors speculate that the cerebellum might enhance the rate of learning about rewards, a possibility supported by the recent discovery in rodents of direct, excitatory projections from the cerebellum to neurons in the brain stem that release the reward-associated neurochemical dopamine¹⁴.

There are several intriguing parallels between the signals found by Sendhilnathan and colleagues and the signals involved in cerebellar control of movement. First, as with

reward-driven learning, for some motor skills, cerebellar Purkinje cells contribute selectively to new motor learning and not to performing older motor skills^{15,16}. Second, Purkinje-cell activity carries information that could guide both ongoing behaviour and the induction of learning during motor- and reward-based learning¹⁷. Third, the Purkinje cells carry signals that could support working memory in the form of activity maintained from one trial to the next in reward-based learning, and in the form of activity maintained during a delay period between a cue and the motor response to the cue, which seems to support motor planning^{11,18}. Finally, during both types of learning, individual Purkinje cells are active for a specific time period of a few hundred milliseconds, with information seemingly passed from cell to cell over time¹⁹. These striking parallels raise the possibility that the cerebellum performs a similar function for error-driven motor learning and reward-driven reinforcement learning.

We learn from both our successes and our failures. These two learning schemes were previously attributed to distinct brain structures, but the current results, along with those of others^{6–12}, blur these mechanistic and conceptual boundaries. As such, the work highlights the need to consider how long-range interactions between brain areas support the shaping of behaviour by experience.

Jennifer L. Raymond is in the Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA. e-mail: jennifer.raymond@stanford.edu

1. Sendhilnathan, N., Ipata, A. E. & Goldberg, M. E. *Neuron* <https://doi.org/10.1016/j.neuron.2019.12.032> (2020).
2. Albus, J. S. *Math. Biosci.* **10**, 25–61 (1971).
3. Schmahmann, J. D. *Neurosci. Lett.* **688**, 62–75 (2019).
4. Rochefort, C., Lefort, J. M., Rondi-Reig, L. *Front. Neural Circuits* **7**, 35 (2013).
5. Wang, S. S., Kloth, A. D. & Badura, A. *Neuron* **83**, 518–532 (2014).
6. Wagner, M. J., Kim, T. H., Savall, J., Schnitzer, M. J. & Luo, L. *Nature* **544**, 96–100 (2017).
7. Heffley, W. *et al. Nature Neurosci.* **21**, 1431–1441 (2018).
8. Heffley, W. & Hull, C. *eLife* **8**, e46764 (2019).
9. Larry, N., Yarkoni, M., Lixenberg, A. & Joshua, M. *eLife* **8**, e46870 (2019).
10. Kostadinov, D., Beau, M., Blanco-Pozo, M. & Häusser, M. *Nature Neurosci.* **22**, 950–962 (2019).
11. Chabrol, F. P., Blot, A. & Masic-Flogel, T. D. *Neuron* **103**, 506–519 (2019).
12. Lixenberg, A., Yarkoni, M., Botschko, Y. & Joshua, M. *J. Neurophysiol.* **123**, 786–799 (2020).
13. Sugrue, L. P., Corrado, G. S. & Newsome, W. T. *Nature Rev. Neurosci.* **6**, 363–375 (2005).
14. Carta, I., Chen, C. H., Schott, A. L., Dorizan, S. & Khodakhah, K. *Science* **363**, eaav0581 (2019).
15. Shutoh, F., Ohki, M., Kitazawa, H., Itoharu, S. & Nagao, S. *Neuroscience* **139**, 767–777 (2006).
16. Jang, D. C., Shim, H. G. & Kim, S. J. Preprint at bioRxiv <https://doi.org/10.1101/513283> (2019).
17. Nguyen-Vu, T. D. B. *et al. Nature Neurosci.* **6**, 1734–1736 (2013).
18. Gao, Z. *et al. Nature* **563**, 113–116 (2018).
19. Li, J. X., Medina, J. F., Frank, L. M. & Lisberger, S. G. *J. Neurosci.* **31**, 12716–12726 (2011).

Coherent electrical control of a single high-spin nucleus in silicon

<https://doi.org/10.1038/s41586-020-2057-7>

Received: 10 June 2019

Accepted: 30 January 2020

Published online: 11 March 2020

 Check for updates

Serwan Asaad^{1,6}, Vincent Mourik^{1,6}, Benjamin Joecker¹, Mark A. I. Johnson¹, Andrew D. Baczewski², Hannes R. Firgau¹, Mateusz T. Mądzik¹, Vivien Schmitt¹, Jarryd J. Pla³, Fay E. Hudson¹, Kohei M. Itoh⁴, Jeffrey C. McCallum⁵, Andrew S. Dzurak¹, Arne Laucht¹ & Andrea Morello^{1,✉}

Nuclear spins are highly coherent quantum objects. In large ensembles, their control and detection via magnetic resonance is widely exploited, for example, in chemistry, medicine, materials science and mining. Nuclear spins also featured in early proposals for solid-state quantum computers¹ and demonstrations of quantum search² and factoring³ algorithms. Scaling up such concepts requires controlling individual nuclei, which can be detected when coupled to an electron^{4–6}. However, the need to address the nuclei via oscillating magnetic fields complicates their integration in multi-spin nanoscale devices, because the field cannot be localized or screened. Control via electric fields would resolve this problem, but previous methods^{7–9} relied on transducing electric signals into magnetic fields via the electron–nuclear hyperfine interaction, which severely affects nuclear coherence. Here we demonstrate the coherent quantum control of a single ¹²³Sb (spin-7/2) nucleus using localized electric fields produced within a silicon nanoelectronic device. The method exploits an idea proposed in 1961¹⁰ but not previously realized experimentally with a single nucleus. Our results are quantitatively supported by a microscopic theoretical model that reveals how the purely electrical modulation of the nuclear electric quadrupole interaction results in coherent nuclear spin transitions that are uniquely addressable owing to lattice strain. The spin dephasing time, 0.1 seconds, is orders of magnitude longer than those obtained by methods that require a coupled electron spin to achieve electrical driving. These results show that high-spin quadrupolar nuclei could be deployed as chaotic models, strain sensors and hybrid spin-mechanical quantum systems using all-electrical controls. Integrating electrically controllable nuclei with quantum dots^{11,12} could pave the way to scalable, nuclear- and electron-spin-based quantum computers in silicon that operate without the need for oscillating magnetic fields.

Nuclear magnetic resonance (NMR) relies on the presence of a static magnetic field, B_0 , that separates the energy levels of the nuclear spins, and a radio-frequency (RF) oscillating magnetic field, B_1 , that induces transitions between such levels. Magnetic fields cannot be easily confined or screened at the nanoscale. Therefore, identical nuclear spins within large regions would all respond to the same signal, preventing the spins from being individually addressed. Electric fields, instead, can be efficiently routed and confined within highly complex nanoscale devices, with a prime example being the sophisticated interconnects found in modern silicon computer chips. These observations suggest that an ideal route to scale up nuclear-spin-based quantum devices would involve the use of RF electric fields for spin control.

A theoretical idea crucial to this strategy was proposed by Bloembergen as early as 1961¹⁰: for nuclei with spin $I > 1/2$ and non-zero electric quadrupole moment q_n , a resonant electric field induces nuclear spin transitions by modulating the nuclear quadrupole interaction, if the nuclei are placed in solids that lack point-inversion symmetry at the lattice site. In bulk ensembles, the static shift of the NMR frequency by a d.c. electric field, named linear quadrupole Stark effect (LQSE), was observed in the 1960s¹³. The resonant version of LQSE, called nuclear electric resonance (NER) was demonstrated only recently¹⁴ in a bulk gallium arsenide (GaAs) crystal.

We report here the demonstration of NER and coherent electrical control of a single antimony (¹²³Sb) nucleus in silicon (Si). The discovery that this nucleus could be electrically controlled was fortuitous. The ¹²³Sb

¹Centre for Quantum Computation and Communication Technology, School of Electrical Engineering and Telecommunications, UNSW Sydney, Sydney, New South Wales, Australia. ²Center for Computing Research, Sandia National Laboratories, Albuquerque, NM, USA. ³School of Electrical Engineering and Telecommunications, UNSW Sydney, Sydney, New South Wales, Australia.

⁴School of Fundamental Science and Technology, Keio University, Yokohama, Japan. ⁵Centre for Quantum Computation and Communication Technology, School of Physics, University of Melbourne, Melbourne, Victoria, Australia. ⁶These authors contributed equally: Serwan Asaad, Vincent Mourik. ✉e-mail: a.morello@unsw.edu.au

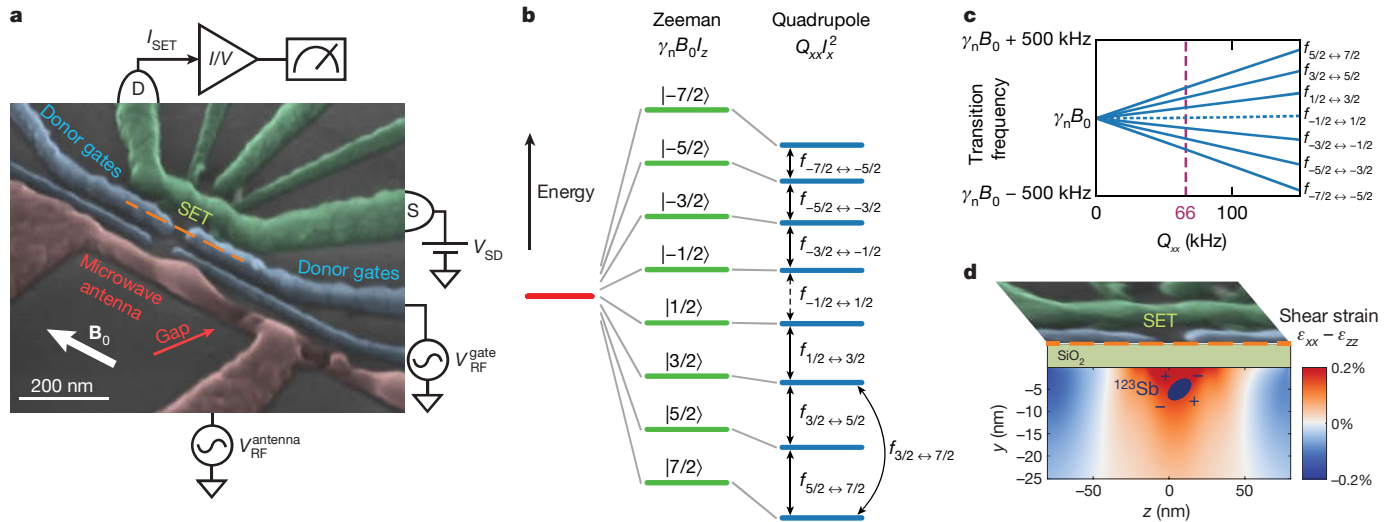


Fig. 1 ^{123}Sb nuclear spin in a silicon device. **a**, False-colour scanning electron micrograph of the silicon metal-oxide-semiconductor device used in the experiment. Note the gaps in the nominally short-circuited antenna terminations. S, source; D, drain; SET, single-electron transistor. **b**, Energy-level diagram of the spin-7/2 nucleus of an ionized ^{123}Sb donor. The magnetic field B_0 introduces a Zeeman splitting (green dashes), and the electric quadrupole interaction Q_{xx} causes a further energy shift (blue dashes). **c**, Nuclear spin

transition frequencies as a function of Q_{xx} . A non-zero Q_{xx} results in seven individually addressable nuclear resonances. The $m_i = -1/2 \leftrightarrow +1/2$ transition (dashed blue line in **c**) is forbidden in NER. The measured quadrupole splitting $f_Q = 66$ kHz is indicated by a dashed purple line. **d**, Shear strain in the silicon substrate, calculated on a vertical cross-section under the orange dashed line in **a**.

donor atom has a nuclear spin of $I = 7/2$ with electric quadrupole moment $q_n = -0.69$ b. Depending on its electrochemical potential relative to a nearby electron reservoir, an electron (with spin $S = 1/2$) may be bound to the nucleus. The atom was implanted in a metal-oxide-semiconductor nanostructure (Fig. 1a) fabricated on isotopically enriched ^{28}Si (for fabrication details, see Supplementary Information section 3), similar to those developed for phosphorus (^{31}P) spin qubits^{5,15,16}. The structure contains a single-electron transistor (SET) for single-shot electron spin readout, which is based on energy-selective electron tunnelling into a cold charge reservoir¹⁷. Four electrostatic gates control the electrochemical potential of the donor, and a broadband on-chip microwave antenna¹⁸ delivers coherent control signals to the donor spins. Single-shot, quantum non-demolition nuclear spin readout⁵ is obtained by combining single-shot electron readout with selective excitation at a specific electron spin resonance frequency, which depends on the nuclear state because of the strong hyperfine interaction (see Methods). The antenna is nominally terminated by a short circuit, in order to obtain maximum current at its tip and produce strong oscillating magnetic fields to control both the electron (at about 40 GHz) and the nuclear (at about 10 MHz) spins of the donor. In this device, however, an electrostatic discharge damaged the short-circuit termination (Fig. 1a). Although the small gap in the termination had a low enough impedance at 40 GHz, to allow current flow for electron spin resonance, at about 10 MHz it produced solely an RF electric field. Once we realized that NER was possible, we began to use the electric gates fabricated exactly above the donor, which had an even stronger effect.

We focus here on the ^{123}Sb donor in its ionized state; the removal of the donor-bound electron precludes any interpretation of the data involving modulation of hyperfine fields^{7,9}. The electron is introduced only for the final readout phase.

In nanoscale Si devices, the aluminium (Al) gates can cause considerable lattice strain at low temperatures, owing to the different thermal contraction of Al and Si (ref. 18). Lattice strain creates an electric field gradient (EFG) of $\mathcal{V}_{\alpha\beta} = \partial^2 V / \partial \alpha \partial \beta$ (V is the electric potential and $\alpha, \beta \in \{x, y, z\}$) at the nuclear site^{20,21} (Fig. 1b), which produces a static nuclear quadrupole interaction $Q_{\alpha\beta} = e q_n \mathcal{V}_{\alpha\beta} / [2I(2I-1)\hbar]$ (\hbar is the Planck

constant and e is the electron charge), resulting in a quadrupole splitting f_Q of the nuclear resonance frequencies (Fig. 1d), making all transitions individually addressable.

The application of an RF electric field of amplitude E_1 modulates the nuclear quadrupole energies by δQ_{xz} and δQ_{yz} , and induces transitions between nuclear states at a rate of $f_{m_i-1 \leftrightarrow m_i}^{\text{Rabi, NER}} \propto |\delta Q_{xz} \langle m_i - 1 | \hat{I}_x \hat{I}_z + \hat{I}_z \hat{I}_x | m_i \rangle|$, where m_i is the secondary spin quantum number, ranging from $-I$ to I in steps of 1, and $\hat{I}_x, \hat{I}_y, \hat{I}_z$ are the eight-dimensional operators describing the x, y, z projections of the $I = 7/2$ spin. Notably, the transition rate is predicted to be zero for the $m_i = -1/2 \leftrightarrow +1/2$ transition (see equation (15) in Supplementary Information section 2C), a consequence of the selection rules of electric quadrupole transitions. Because the quadrupole interaction is quadratic in the nuclear spin operators, first-order transitions between spin states that differ by $\Delta m_i = \pm 2$ are allowed. These occur at a rate of $f_{m_i-2 \leftrightarrow m_i}^{\text{Rabi, NER}} \propto |\delta Q_{xx} \langle m_i - 2 | \hat{I}_x^2 | m_i \rangle|$ (see equation (19) in Supplementary Information section 2C) and, importantly, all $\Delta m_i = \pm 2$ transitions have a non-zero rate.

Figure 2a shows the experimental NER spectrum for $\Delta m_i = \pm 1$ transitions, which contains six sharp resonances separated by $f_Q = 66$ kHz. The $m_i = -1/2 \leftrightarrow +1/2$ transition is absent, as expected from NER. All six predicted $\Delta m_i = \pm 2$ transitions are observed (Fig. 2b). The ability to excite the $m_i = -1/2 \leftrightarrow +3/2$ transition was used to ‘jump over’ the forbidden $m_i = -1/2 \leftrightarrow +1/2$ transition and observe the $\Delta m_i = \pm 1$ transitions at negative m_i , which would otherwise be inaccessible if starting from a positive m_i . Similarly, the NER spectrum for $\Delta m_i = \pm 2$ transitions (Fig. 2b) could be completed only by employing a $\Delta m_i = \pm 1$ transition.

Figure 2c, d presents the observed transition rates between each pair of states, in excellent agreement with the predicted trends from NER theory. Using NMR, the Rabi frequencies for the $\Delta m_i = \pm 1$ transitions would be $f_{m_i-1 \leftrightarrow m_i}^{\text{Rabi, NMR}} \propto |\gamma_n B_1 \langle m_i - 1 | \hat{I}_x | m_i \rangle|$ ($\gamma_n = 5.55 \text{ MHz T}^{-1}$ is the nuclear gyromagnetic ratio), which is notably maximal for the $m_i = -1/2 \leftrightarrow +1/2$ transition. The $\Delta m_i = \pm 2$ NMR transitions are forbidden to first-order. These results prove decisively that our experiments do not constitute a form of magnetic resonance.

As observed in earlier experiments on ^{31}P (refs. 16,22), the nuclear spins of ionized donors in ^{28}Si have exceptional quantum coherence

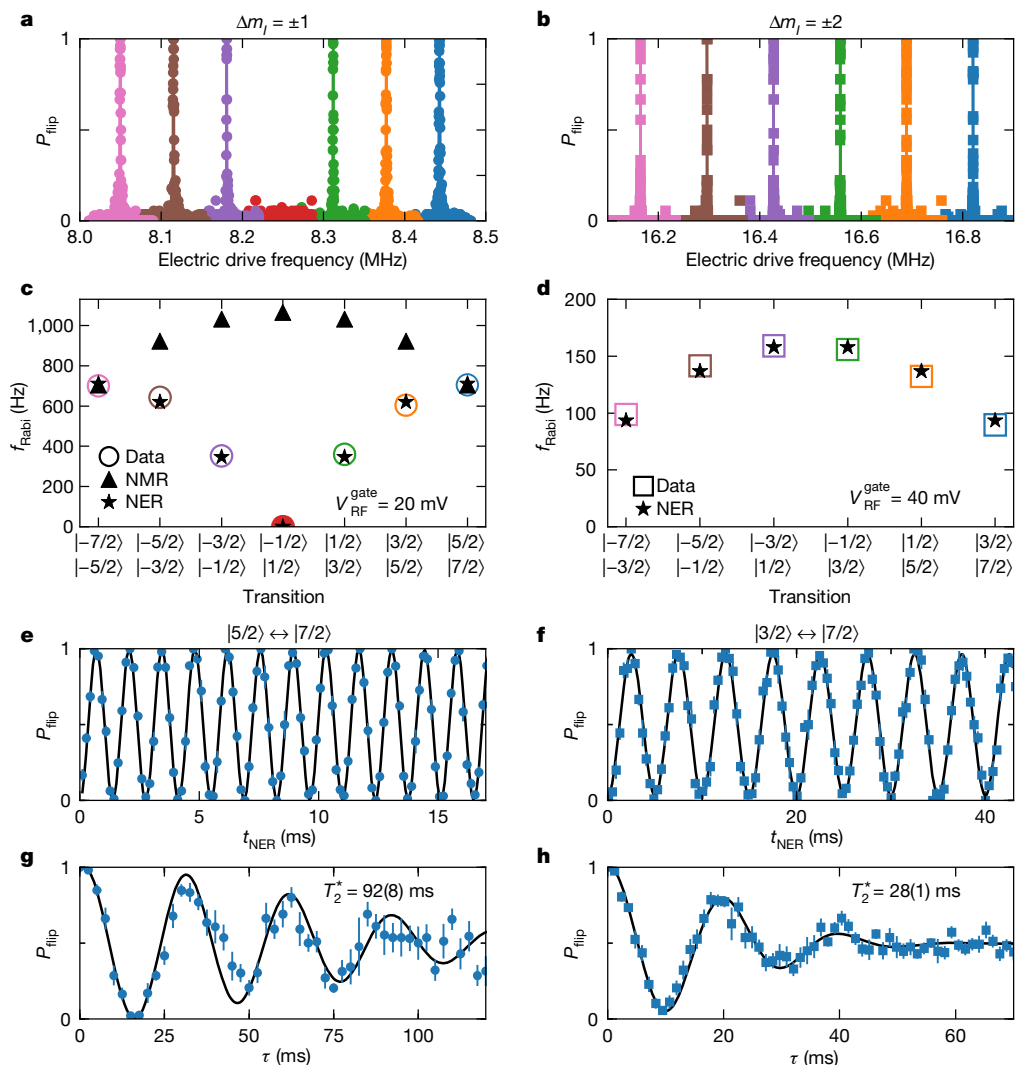


Fig. 2 | Nuclear electric resonance. **a, b**, NER spectrum for the $\Delta m_I = \pm 1$ (**a**) and $\Delta m_I = \pm 2$ (**b**) transitions, obtained by applying voltage $V_{\text{RF}}^{\text{gate}}$ to a donor gate (see Fig. 1a). The $m_I = -1/2 \leftrightarrow +1/2$ transition (**a**, red) was not observed, as expected in NER. To acquire the complete $\Delta m_I = \pm 1$ spectrum, the $m_I = -1/2 \leftrightarrow +3/2$ transition was used to bridge the positive and negative m_I values. P_{flip} represents the probability of flipping the nuclear spin between two states. **c, d**, Rabi frequencies of the $\Delta m_I = \pm 1$ (**c**) and $\Delta m_I = \pm 2$ (**d**) transitions, each measured at a constant NER drive amplitude (see Extended Data Fig. 2 for the corresponding Rabi oscillations). Measured values (circles in **c**, squares in **d**) are compared to the theoretical predictions for NER (stars) and NMR (triangles in **c**, and are incompatible with NMR. **e, f**, Nuclear Rabi oscillations on the $m_I = +5/2 \leftrightarrow +7/2$ (**e**) and $m_I = +3/2 \leftrightarrow +7/2$ (**f**) transitions. A sinusoid with no decay is used to fit the data. t_{NER} , NER pulse duration. **g, h**, Nuclear Ramsey fringes used to extract the pure dephasing time T_{2n}^* on the $m_I = +5/2 \leftrightarrow +7/2$ (**g**) and $m_I = +3/2 \leftrightarrow +7/2$ (**h**) transitions. The fits are sinusoids with envelopes decaying as $\exp[-(\tau/T_{2n}^*)^2]$, where τ is the free precession time. Error bars and uncertainties denote the 68% confidence level.

properties. We performed a Ramsey experiment (Fig. 2g) on the $m_I = +5/2 \leftrightarrow +7/2$ ($\Delta m_I = \pm 1$) transition to extract the pure dephasing time T_{2n}^* ($+5/2 \leftrightarrow +7/2$) = 92(8) ms (68% confidence level), which corresponds to an NER broadening (full-width at half-maximum) of $\Gamma_n = \ln 2 / (\pi T_{2n}^*) = 2.4(2)$ Hz. The $m_I = +3/2 \leftrightarrow +7/2$ ($\Delta m_I = \pm 2$) transition has shorter dephasing time, T_{2n}^* ($+3/2 \leftrightarrow +7/2$) = 28(1) ms (Fig. 2h). Both values, although extremely long in absolute terms, are noticeably shorter than the time $T_{2n}^* = 250$ –600 ms measured on the ^{31}P nucleus in two other similar devices¹⁶ fabricated on the same ^{28}Si wafer. Given that the ^{31}P nucleus has zero quadrupole moment, this suggests that the ^{123}Sb coherence may be affected by electrical noise²³, in a way that the ^{31}P coherence is not. Nonetheless, our dephasing time remains two orders of magnitude longer than that observed in ^{31}P when adding a hyperfine coupled electron, $T_{2n}^* \approx 430$ –570 μs (ref. ¹⁶) and three orders of magnitude longer than the observed $T_2^* = 64$ μs of a terbium nucleus in a single-atom magnet⁷. This observation highlights the benefit of a purely electrical control mechanism that does not rely on hyperfine interactions.

We measured the Rabi frequencies of the $\Delta m_I = \pm 1$ and $\Delta m_I = \pm 2$ NER transitions as a function of the amplitude of the RF voltage applied to the gate, finding transition rates $g_{E,1} = 34.21(3)$ Hz mV^{-1} (Fig. 3a) and $g_{E,2} = 1.995(4)$ Hz mV^{-1} (Fig. 3b). These transition rates show that NER is a weak effect but, owing to the long nuclear spin coherence in ^{28}Si , we were able to perform high-fidelity Rabi flops persisting for tens of milliseconds (Fig. 2e, f).

In addition to driving nuclear spin transitions with an RF voltage, we were able to apply Stark shifts to the resonance frequencies using an

additional d.c. voltage $\Delta V_{\text{DC}}^{\text{gate}}$ on the gates (Fig. 3c, d). All NER frequencies shifted according to $\Delta f = (\partial f_0 / \partial V_{\text{DC}}^{\text{gate}}) [\Delta m_I] [m_I - (\Delta m_I/2)] \Delta V_{\text{DC}}^{\text{gate}}$, where $\partial f_0 / \partial V_{\text{DC}}^{\text{gate}} = 9.9(3)$ Hz mV^{-1} , and $[\Delta m_I] [m_I - (\Delta m_I/2)]$ is a factor of order unity that represents the matrix element of the electric quadrupole interaction between the initial and final state of each transition (see Supplementary Information section 2C for details).

The results reported here constitute the first, to our knowledge, observation of coherent, purely electrical control of a single nuclear spin. Achieving this in silicon is, at first sight, remarkable: no effect of electric fields on nuclear spins has ever been observed in a non-polar, non-piezoelectric material in the absence of a hyperfine-coupled electron. To gain a microscopic understanding of this phenomenon, we conjectured that our results are a form of LQSE¹³. Resonant transitions between nuclear levels induced by electric fields (NER) require that the crystal does not possess point-inversion symmetry at the atomic site¹⁰, as is indeed the case for silicon. The observation of individual NER transitions, separated by the nuclear quadrupole splitting f_0 , implies that a static EFG must exist at the nuclear site. This requires breaking the T_d (tetrahedral) symmetry of the silicon crystal, as it would otherwise have zero net EFG. For instance, uniaxial strain (for example, ϵ_{zz}) lowers the symmetry to D_{2d} (tetragonal scalenohedral), whereas shear strain (for example, $\epsilon_{xx} - \epsilon_{zz}$) lowers it to C_{2v} (rhombohedral pyramidal). The T_d symmetry can also be broken by an electric field that polarizes the atomic bonds. This latter effect explains both the observation of NER and the static shift of the nuclear spin resonance lines (Fig. 3c, d) due to LQSE on application of a static gate voltage.

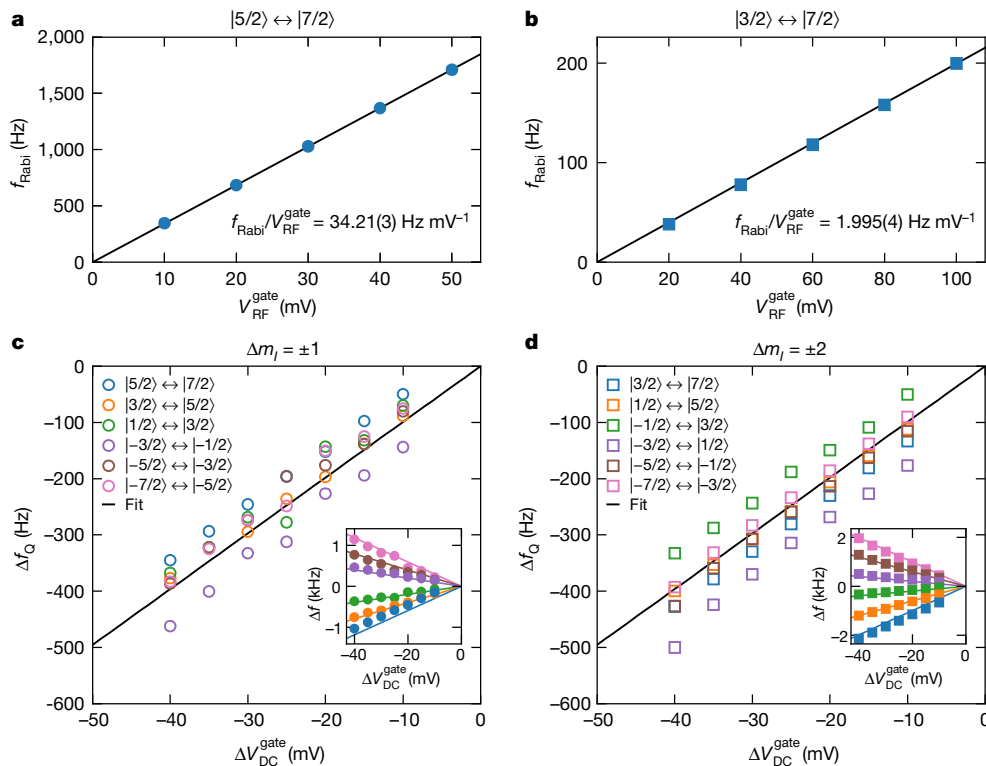


Fig. 3 | Linear quadrupole Stark effect. **a, b**, Rabi frequencies f_{Rabi} for varying electric drive peak amplitude $V_{\text{RF}}^{\text{gate}}$, measured on the $\Delta m_I = \pm 1$ transition $|5/2\rangle \leftrightarrow |7/2\rangle$ (**a**) and the $\Delta m_I = \pm 2$ transition $|3/2\rangle \leftrightarrow |7/2\rangle$ (**b**). The linear relationship between $V_{\text{RF}}^{\text{gate}}$ and f_{Rabi} is consistent with a first-order transition induced by the LQSE. **c, d**, Quadrupole shift $\Delta f_Q = (\partial f_Q / \partial V_{\text{DC}}^{\text{gate}}) \Delta V_{\text{DC}}^{\text{gate}}$ measured while applying an additional d.c. voltage $\Delta V_{\text{DC}}^{\text{gate}}$ on a donor gate. The application of $\Delta V_{\text{DC}}^{\text{gate}}$ causes each transition frequency $f_{m_I \leftrightarrow m_I'}$ to shift by $\Delta f = (\partial f_Q / \partial V_{\text{DC}}^{\text{gate}}) [\Delta m_I (m_I - (\Delta m_I/2))] \Delta V_{\text{DC}}^{\text{gate}}$ (inset; see Extended Data Fig. 3 for nuclear spectra). A combined fit through all $\Delta m_I = \pm 1$ (**c**) and $\Delta m_I = \pm 2$ (**d**) frequency shifts results in an LQSE coefficient of $\partial f_Q / \partial V_{\text{DC}}^{\text{gate}} = 9.9(3) \text{ Hz mV}^{-1}$.

The larger and charged donor atom introduces a local lattice distortion, displacing its four coordinating Si atoms by 0.2 Å, and polarizes the charge density along the bonds (Fig. 4b, d). This, however, does not yet break the T_d symmetry. An EFG is obtained by further introducing strain. The S tensor that links EFG to strain has two unique components, S_{11} (uniaxial) and S_{44} (shear). We conducted a first-principles, density functional theory calculation and extracted $S_{11} = 2.4 \times 10^{22} \text{ V m}^{-2}$ and $S_{44} = 6.1 \times 10^{22} \text{ V m}^{-2}$ (see Supplementary Information section 7C2 for details). Using a finite-element numerical model we computed the strain profile in our device, as caused by the different thermal expansions of Si and Al on cooling to cryogenic temperatures^{19,21} (Fig. 1b). Finally, we triangulated the most likely location of the ^{123}Sb donor by combining the implantation depth profile with a model of the relative capacitive coupling between the donor and different pairs of control gates, extracted from the experimental charge stability diagrams (see Extended Data Fig. 4 and Supplementary Information section 7C3 for details). By combining these three pieces of information, we arrived at a spatial map of quadrupole splittings f_Q (Fig. 4c), which shows good agreement between the models and the experiment around the predicted location of the donor under study.

The effect of electric fields on the quadrupole interaction, both static (LQSE) and dynamic (NER), can be understood as arising from the single unique component of the R tensor, R_{14} (see Supplementary Information section 7D1 for details). By combining a finite-element model of the electric field in the device, the estimated $^{123}\text{Sb}^+$ donor position and the experimental values of LQSE and NER Rabi frequencies, we extracted $R_{14} = 1.7 \times 10^{12} \text{ m}^{-1}$ (see Supplementary Information section 7D2 for details). The strength of this coupling is comparable to prior bulk measurements of LQSE on arsenic (^{75}As) in GaAs (ref. 24). This can be understood by observing that, although the Sb^+-Si bond has a weaker ionic character than the $\text{Ga}-\text{As}$ bond, R_{14} scales with atomic number, leading to a similar overall value. Given that our model agrees with the experiment within a factor of order unity and no alternative explanation comes within orders of magnitude of the results (see Supplementary Information section 7E for details), we conclude that we

have observed the manifestation of LQSE and NER in a single nuclear spin in silicon.

Our results have substantial consequences for the development of nuclear-spin-based quantum computers and the design of nanoscale quantum devices. The Hilbert space of the $I = 7/2$ ^{123}Sb nucleus has eight dimensions. It can encode the equivalent of three quantum bits of information, allowing simple quantum algorithms²⁵ or quantum error correction codes²⁶, all using solely electric fields. The donor electron and nuclear spins combined form a ‘flip-flop’ qubit¹¹, controllable by electric-dipole spin resonance. This scheme normally requires a magnetic antenna to reset the nuclear state in the appropriate qubit subspace. This need could be removed completely by using an electrically drivable high-spin nucleus such as ^{123}Sb . A recent result showed that lithographic quantum dots in silicon can be entangled with nuclear spins and that the nuclear coherence can be preserved while shuttling the electron between different dots¹². Electron spin qubits in silicon can be coherently controlled by electric fields with high speed and high fidelity²⁷. Adding the ability to electrically control quadrupolar nuclei paves the way to quantum computer architectures that integrate fast electron spin qubits with long-lived nuclear quantum memories while fully exploiting the controllability and scalability of silicon metal-oxide-semiconductor devices, without the complication of routing RF magnetic fields within the device.

The experimental validation of a microscopic model of the relation between strain and quadrupole splitting, obtained in a functional silicon electronic device, suggests the use of quadrupolar nuclei as single-atom probes of local strain, which has a key role in enhancing the performance of ultra-scaled transistors²⁸.

The NER methods and microscopic models presented here could be extended to the study of quadrupolar nuclei in materials such as diamond and silicon carbide, where electrical and strain tuning of optically addressable electronic spins has been demonstrated^{29,30}.

The observation of a large quadrupole splitting of $f_Q = 66 \text{ kHz}$ in a high-spin nucleus creates a platform in which to study quantum chaotic dynamics in a single particle³¹. This has further applications in quantum

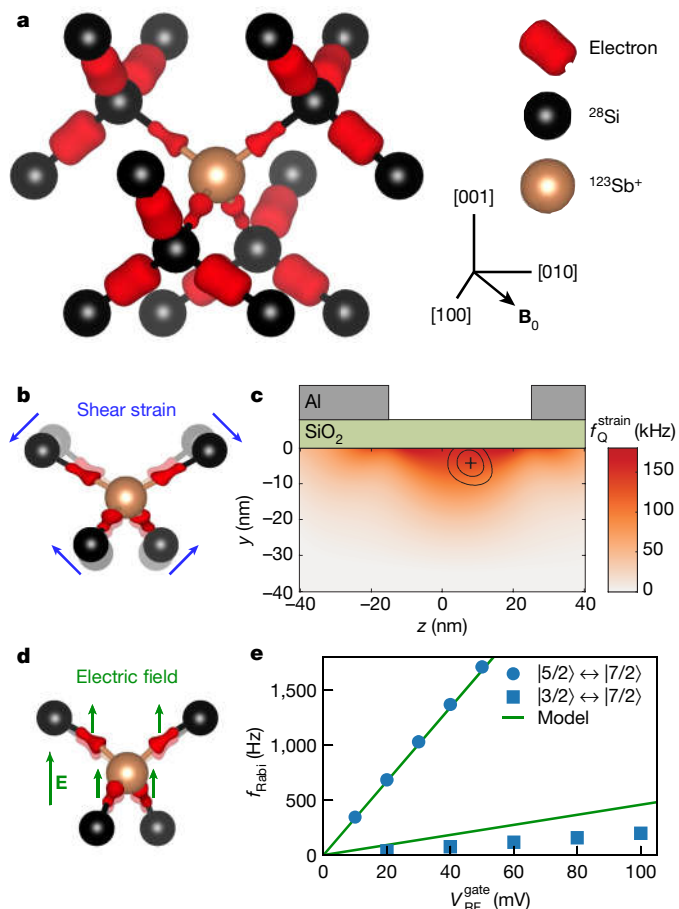


Fig. 4 | Microscopic origins of the quadrupole interaction. **a**, Valence charge density near the Sb^+ atom (gold) and its 16 closest Si atoms (black) with a charge density isosurface (red). The positive charge of the donor causes an asymmetric charge density along the Sb^+ –Si bond but, in the absence of strain or external electric fields, the EFG at the ^{123}Sb site vanishes by symmetry. **b**, Shear strain displaces the Si atoms and covalent bonds neighbouring the ^{123}Sb nucleus, creating an EFG that results in a quadrupole shift. **c**, Quadrupole splitting f_Q , predicted by combining density functional theory calculations and finite-element simulations (see Supplementary Information section 7C for details). Black contours enclose the 68% and 95% confidence regions for the location of the donor, as obtained from capacitance triangulation and the donor implantation profile (see Supplementary Information section 7A for details). **d**, Electric fields applied via the gate voltage distort the charge distribution, resulting in both linear frequency shifts (LQSE) and coherent spin transitions (NER). **e**, Calculation of the NER Rabi frequencies caused by electrical EFG modulation (green lines), compared to experimental results for a $\Delta m_I = \pm 1$ (dots) and a $\Delta m_I = \pm 2$ (squares) transition. All f_{Rabi} values are determined using a single parameter, R_{14} , calculated via finite-element modelling and electronic structure theory. No free fitting parameters were used.

information science, for example, because of the remarkable analogies between chaotic spin models and digital quantum simulations³².

Although the strain in the present device is static, our work allows us to predict the nuclear Rabi frequencies that would arise from time-dependent strain (see Supplementary Information section 8 for details). A dynamical strain of about 5×10^{-8} would cause a Rabi frequency of 10 Hz, comparable to both the inhomogeneous nuclear linewidth $\Gamma_n \approx 2.4$ Hz and to the linewidth Γ_n of high-quality silicon mechanical resonators in the megahertz range³³. Therefore, it is conceivable that the strong-coupling limit of cavity quantum electrodynamics might be achieved between a single nuclear spin and a macroscopic mechanical oscillator, adding a novel spin–mechanical coupling pathway to the

toolbox of hybrid quantum systems for quantum information processing and precision sensing³⁴.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2057-7>.

- Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
- Jones, J. A., Mosca, M. & Hansen, R. H. Implementation of quantum search algorithm on a quantum computer. *Nature* **393**, 344–346 (1998).
- Vandersypen, L. M. K. et al. Experimental realization of Shor's factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
- Jelezko, F. et al. Observation of coherent oscillation of a single nuclear spin and realization of a two-qubit conditional quantum gate. *Phys. Rev. Lett.* **93**, 130501 (2004).
- Pla, J. J. et al. High-fidelity readout and control of a nuclear spin qubit in silicon. *Nature* **496**, 334–338 (2013).
- Willke, P. et al. Hyperfine interaction of individual atoms on a surface. *Science* **362**, 336–339 (2018).
- Thiele, S. et al. Electrically driven nuclear spin resonance in single-molecule magnets. *Science* **344**, 1135–1138 (2014).
- Laucht, A. et al. Electrically controlling single-spin qubits in a continuous microwave field. *Sci. Adv.* **1**, e1500022 (2015).
- Sigillito, A. J., Tyryshkin, A. M., Schenkel, T., Houck, A. A. & Lyon, S. A. All-electric control of donor nuclear spin qubits in silicon. *Nat. Nanotechnol.* **12**, 958–962 (2017).
- Bloembergen, N. Linear Stark effect in magnetic resonance spectra. *Science* **133**, 1363–1364 (1961).
- Tosi, G. et al. Silicon quantum processor with robust long-distance qubit couplings. *Nat. Commun.* **8**, 450 (2017).
- Hensen, B. et al. A silicon quantum-dot-coupled nuclear spin qubit. *Nat. Nanotechnol.* **15**, 13–17 (2020).
- Dixon, R. & Bloembergen, N. Electrically induced perturbations of halogen nuclear quadrupole interactions in polycrystalline compounds. ii. Microscopic theory. *J. Chem. Phys.* **41**, 1739–1747 (1964).
- Ono, M., Ishihara, J., Sato, G., Ohno, Y. & Ohno, H. Coherent manipulation of nuclear spins in semiconductors with an electric field. *Appl. Phys. Express* **6**, 033002 (2013).
- Pla, J. J. et al. A single-atom electron spin qubit in silicon. *Nature* **489**, 541–545 (2012).
- Muhonen, J. T. et al. Storing quantum information for 30 seconds in a nanoelectronic device. *Nat. Nanotechnol.* **9**, 986–991 (2014).
- Morello, A. et al. Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).
- Dehollain, J. et al. Nanoscale broadband transmission lines for spin qubit control. *Nanotechnology* **24**, 015202 (2013).
- Thorbeck, T. & Zimmerman, N. M. Formation of strain-induced quantum dots in gated semiconductor nanostructures. *AIP Adv.* **5**, 087107 (2015).
- Franke, D. P. et al. Interaction of strain and nuclear spins in silicon: quadrupolar effects on ionized donors. *Phys. Rev. Lett.* **115**, 057601 (2015).
- Pla, J. J. et al. Strain-induced spin-resonance shifts in silicon devices. *Phys. Rev. Appl.* **9**, 044014 (2018).
- Saeedi, K. et al. Room-temperature quantum bit storage exceeding 39 minutes using ionized donors in silicon-28. *Science* **342**, 830–833 (2013).
- Franke, D. P., Pflüger, M. P. D., Itoh, K. M. & Brandt, M. S. Multiple-quantum transitions and charge-induced decoherence of donor nuclear spins in silicon. *Phys. Rev. Lett.* **118**, 246401 (2017).
- Gill, D. & Bloembergen, N. Linear Stark splitting of nuclear spin levels in GaAs. *Phys. Rev.* **129**, 2398–2403 (1963).
- Godfrin, C. et al. Operating quantum states in single magnetic molecules: implementation of Grover's quantum algorithm. *Phys. Rev. Lett.* **119**, 187702 (2017).
- Waldherr, G. et al. Quantum error correction in a solid-state hybrid spin register. *Nature* **506**, 204–207 (2014).
- Yoneda, J. et al. A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102–106 (2018).
- Thompson, S. E., Sun, G., Choi, Y. S. & Nishida, T. Uniaxial-process-induced strained-Si: extending the CMOS roadmap. *IEEE Trans. Electron Dev.* **53**, 1010–1020 (2006).
- Dolde, F. et al. Electric-field sensing using single diamond spins. *Nat. Phys.* **7**, 459–463 (2011).
- Falk, A. L. et al. Electrically and mechanically tunable electron spins in silicon carbide color centers. *Phys. Rev. Lett.* **112**, 187601 (2014).
- Mourik, V. et al. Exploring quantum chaos with a single nuclear spin. *Phys. Rev. E* **98**, 042206 (2018).
- Sieberer, L. M. et al. Digital quantum simulation, trotter errors, and quantum chaos of the kicked top. *npj Quantum Inf.* **5**, 78 (2019).
- Ghaffari, S. et al. Quantum limit of quality factor in silicon micro and nano mechanical resonators. *Sci. Rep.* **3**, 3244 (2013); corrigendum 4, 4331 (2013).
- Kurizki, G. et al. Quantum technologies with hybrid systems. *Proc. Natl Acad. Sci. USA* **112**, 3866–3873 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

Fabrication

The device was fabricated on a (100) p-type silicon wafer, with a 900-nm-thick epitaxial layer of isotopically purified ^{28}Si on top (concentration of residual ^{29}Si , 730 ppm). Metallic leads for the SET were formed using optical lithography and phosphorus diffusion. The substrate was subsequently covered with a 200-nm-thick field oxide, with a small central window ($10 \times 20 \mu\text{m}^2$) containing a high-quality, thermally grown layer of SiO_2 with a thickness of 8 nm. Using a combination of standard optical and electron-beam lithography techniques, the device was fabricated on this thin oxide window. First, a small ($90 \times 100 \text{ nm}^2$) window was defined, through which ^{123}Sb ions were implanted at an energy of 8 keV and a fluence of $2 \times 10^{11} \text{ cm}^{-2}$, corresponding to an average of 14 donors in the implantation window. Donors were activated using a rapid thermal anneal at 1,000 °C for 5 s. Next, in two electron-beam lithography steps, the gates forming the SET, the donor gates and the microwave antenna were created using thermally evaporated aluminium and lift-off, with native aluminium oxide as the gate dielectric. Ohmic contacts to the n-doped SET leads were formed using optical lithography, evaporated Al and lift-off, followed by a forming gas anneal. A detailed step-by-step process flow is given in Supplementary Information section 3.

Experimental setup

The sample was cooled to a temperature of 20 mK in a dilution refrigerator (Bluefors BF-LD400) fitted with a superconducting magnet. During the measurements, arbitrary waveform generators (Signadyne M3201A and M3300A) were used to tune the donor electrochemical potential, generate NER pulses and IQ-modulate the microwave signals generated by a vector microwave source (Keysight E8267D). The SET current was amplified with a transimpedance amplifier (FEMTO DLPCA-200 in combination with Stanford Instruments SIM911) and subsequently measured with a digitizer (Signadyne M3300A). Full details of the experimental setup, including a wiring schematic, can be found in Supplementary Information section 4.

Nuclear spin readout

The nuclear spin state is measured via electron spin readout. For nuclear spin readout, an electron is introduced to the donor by tuning its electrochemical potential about the Fermi level of the SET such that a spin-down electron tunnels onto the donor. The electron spin resonance (ESR) spectrum (Extended Data Fig. 1) shows eight distinct resonance lines, each corresponding to a single nuclear spin eigenstate. The electron spin can be flipped conditionally on the nuclear spin state, resulting in single-shot nuclear spin readout. Electron spin readout is achieved by spin-to-charge conversion through spin-dependent tunnelling onto an SET and subsequent detection of the change in charge occupation of the donor (see Supplementary Information section 5 for details). As each of these electron spin measurements project the nucleus into a single spin eigenstate, this is a quantum non-demolition measurement. Therefore, each single-shot nuclear spin readout can be repeated to increase the nuclear spin readout fidelity while retaining the single-shot nature of the nuclear spin measurement. An NER pulse has a probability P_{flip} of flipping the nuclear spin between two states. To measure P_{flip} , an NER pulse followed by nuclear spin readout is performed $N_{\text{iterations}}$ times. The first record of the nuclear spin state is used as a reference, and each subsequent record is compared to the one before it. This reveals the number of times that the nucleus flips, N_{flips} , between the two spin states. Therefore, the flip probability is simply the number of flips per number of recorded attempts, that is, $P_{\text{flip}} = N_{\text{flips}} / (N_{\text{iterations}} - 1)$.

Theoretical modelling

The spin Hamiltonian of the ^{123}Sb nucleus takes the form:

$$\frac{\hat{H}}{h} = \gamma_n B_0 \hat{I}_z + \sum_{\alpha, \beta \in \{x, y, z\}} Q_{\alpha\beta} \hat{I}_\alpha \hat{I}_\beta \quad (1)$$

where $h = 6.626 \times 10^{-34} \text{ J Hz}^{-1}$ is the Planck constant, $\gamma_n = -5.553 \text{ MHz T}^{-1}$ is the nuclear gyromagnetic ratio and $B_0 = 1.496 \text{ T}$. In the presence of an RF electric field of amplitude E_1 , the $\Delta m_I = \pm 1$ transitions are driven by an additional Hamiltonian term of the form:

$$\hat{H}_{m_I-1 \leftrightarrow m_I}^{\text{NER}}(t)/h = \cos(2\pi f t) [\delta Q_{xz} (\hat{I}_x \hat{I}_z + \hat{I}_z \hat{I}_x) + \delta Q_{yz} (\hat{I}_y \hat{I}_z + \hat{I}_z \hat{I}_y)] \quad (2)$$

The $\Delta m_I = \pm 2$ transitions are driven by a term of the form:

$$\hat{H}_{m_I-2 \leftrightarrow m_I}^{\text{NER}}(t)/h = \cos(2\pi f t) [\delta Q_{xx} \hat{I}_x^2 + \delta Q_{yy} \hat{I}_y^2 + \delta Q_{xy} (\hat{I}_x \hat{I}_y + \hat{I}_y \hat{I}_x)] \quad (3)$$

A detailed derivation of the matrix elements responsible for driving the $\Delta m_I = \pm 1$ and $\Delta m_I = \pm 2$ NER transitions is given in Supplementary Information section 2C. A finite-element model is used to compute the strain and electric fields in the silicon layer near the donor position using the COMSOL multiphysics software. The donor position is triangulated by comparing simulated gate-to-donor coupling strengths with the experimentally observed strength, combined with the donor implantation profile (see Extended Data Fig. 4 and Supplementary Information section 7A). Kohn–Sham density functional theory is employed to calculate the components of the S tensor that describe the impact of strain on the EFG. To this end, 64- and 512-atom supercells were strained using the PAW (projector augmented-wave) formalism³⁵ with a plane-wave basis, as implemented in VASP (Vienna ab initio simulation package)^{36–38}. The electric-field response tensor is estimated by comparing the data points from the d.c. LQSE (Fig. 3c, d) and $\Delta m_I = \pm 1$ (Fig. 2c) and $\Delta m_I = \pm 2$ (Fig. 2d) Rabi frequencies with the simulated electric fields at the triangulated donor position. The final R_{14} is found by minimizing the normalized residuals of the three separate R_{14} estimates. Full theoretical modelling details can be found in Supplementary Information section 7.

Data availability

All data necessary to evaluate the claims of this paper are provided in the main manuscript and Supplementary Information. Raw data files, data analysis code and simulation code are available at <https://doi.org/10.26190/5de9c295a8821>.

35. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
36. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
37. Kresse, G. & Furthmüller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
38. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
39. Mansir, J. et al. Linear hyperfine tuning of donor spins in silicon using hydrostatic strain. *Phys. Rev. Lett.* **120**, 167701 (2018).

Acknowledgements We thank T. Botzem and J. T. Muhonen for discussions. The research was funded by the Australian Research Council Discovery Projects (grants DP150101863 and DP180100969) and the Australian Department of Industry, Innovation and Science (grant AUMURIO0002). V.M. acknowledges support from a Niels Stensen Fellowship. M.A.I.J. and H.R.F. acknowledge the support of Australian Government Research Training Program Scholarships. J.J.P. is supported by an Australian Research Council Discovery Early Career Research Award (DE190101397). A.M. was supported by a Weston Visiting Professorship at the Weizmann Institute of Science during part of the writing of this manuscript. We acknowledge support from the Australian National Fabrication Facility (ANFF), and from the laboratory of R. Elliman at the Australian National University for the ion implantation facilities. A.D.B. was supported by the Laboratory Directed Research and Development programme at Sandia National Laboratories, Project 213048. Sandia National Laboratories is a multi-missions laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the National Nuclear Security Administration of the US Department of Energy under contract DE-NA0003525. The views expressed in this manuscript do not necessarily represent the views of the US Department of Energy or the US Government. K.M.I. acknowledges support from Grant-in-Aid for Scientific Research by MEXT.

Author contributions S.A. and M.A.I.J. performed the measurements under the supervision of V.M., A.L. and A.M., with the assistance of V.S., M.T.M. and H.R.F.; S.A. and M.A.I.J. analysed the data under the supervision of V.M. and A.M., with the assistance of H.R.F., V.S., J.J.P. and A.L.; A.D.B., S.A., V.M., B.J. and A.M. developed a microscopic theory supported by finite-element modelling by B.J. and electronic structure calculations by A.D.B.; F.E.H. partially fabricated the device under the supervision of A.S.D., on isotopically enriched material supplied by K.M.I. and M.T.M. subsequently fabricated the aluminium gate structures under the supervision of V.M. and A.M.; J.C.M. designed and performed the ^{123}Sb ion implantation; S.A., V.M., B.J., M.A.I.J., A.D.B., H.R.F. and A.M. wrote the manuscript and Supplementary Information, with input from all co-authors; A.M. initiated and supervised the research programme.

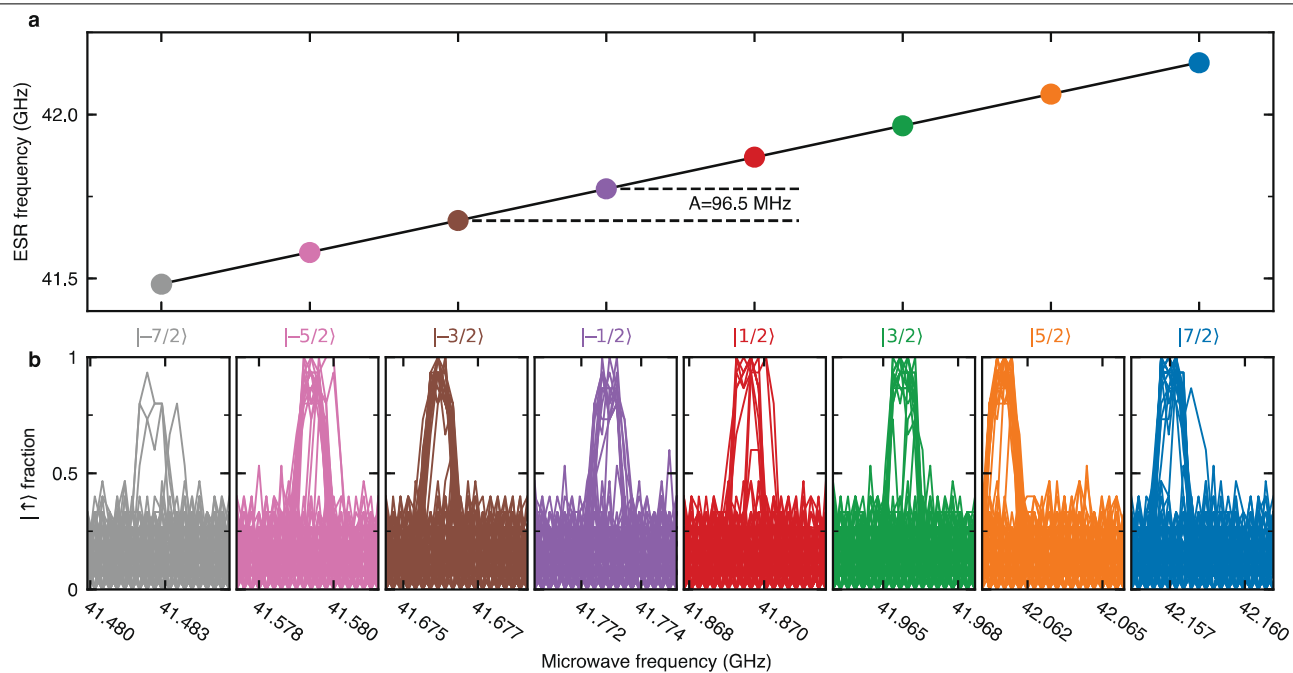
Competing interests S.A., V.M. and A.M. have submitted a patent application that describes the use of electrically controlled high-spin nuclei for quantum information processing (AU2018900665A).

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2057-7>.

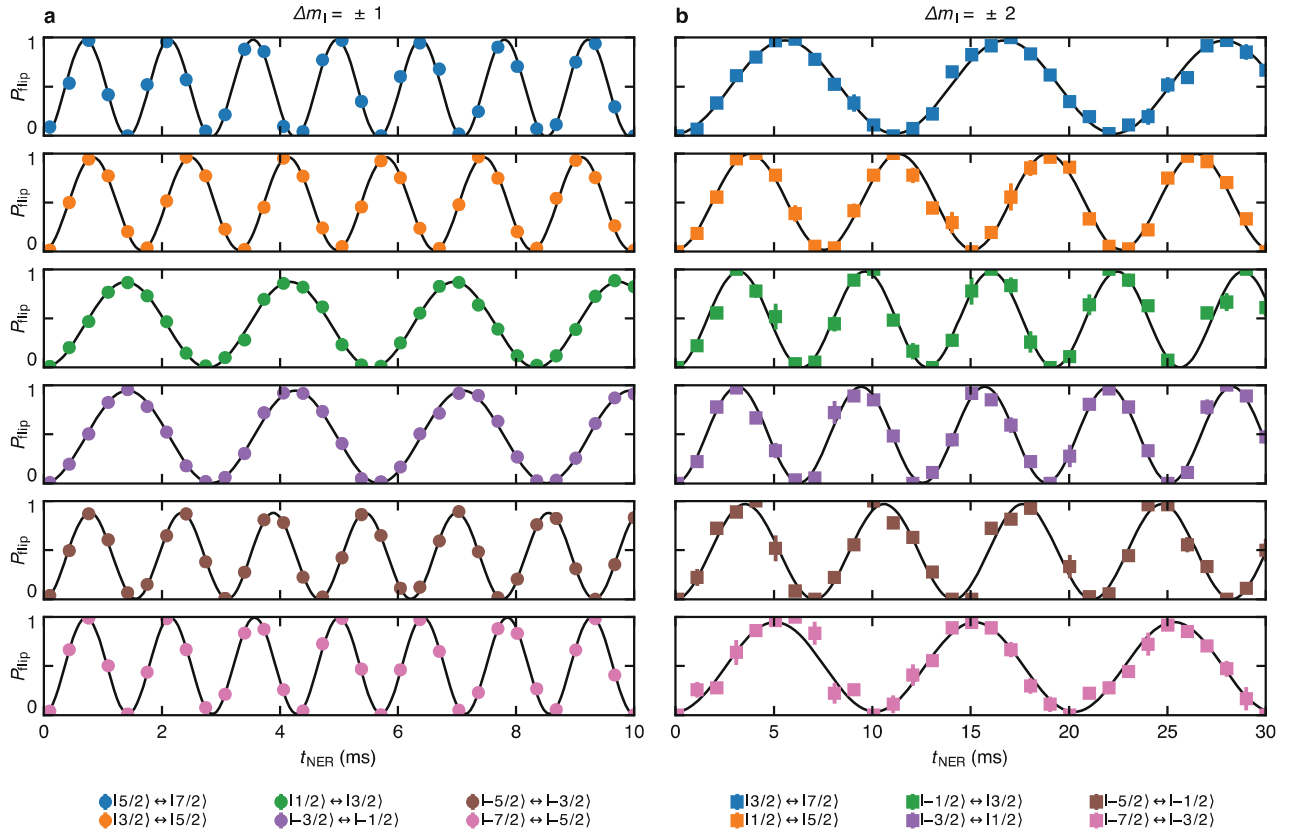
Correspondence and requests for materials should be addressed to A.M.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



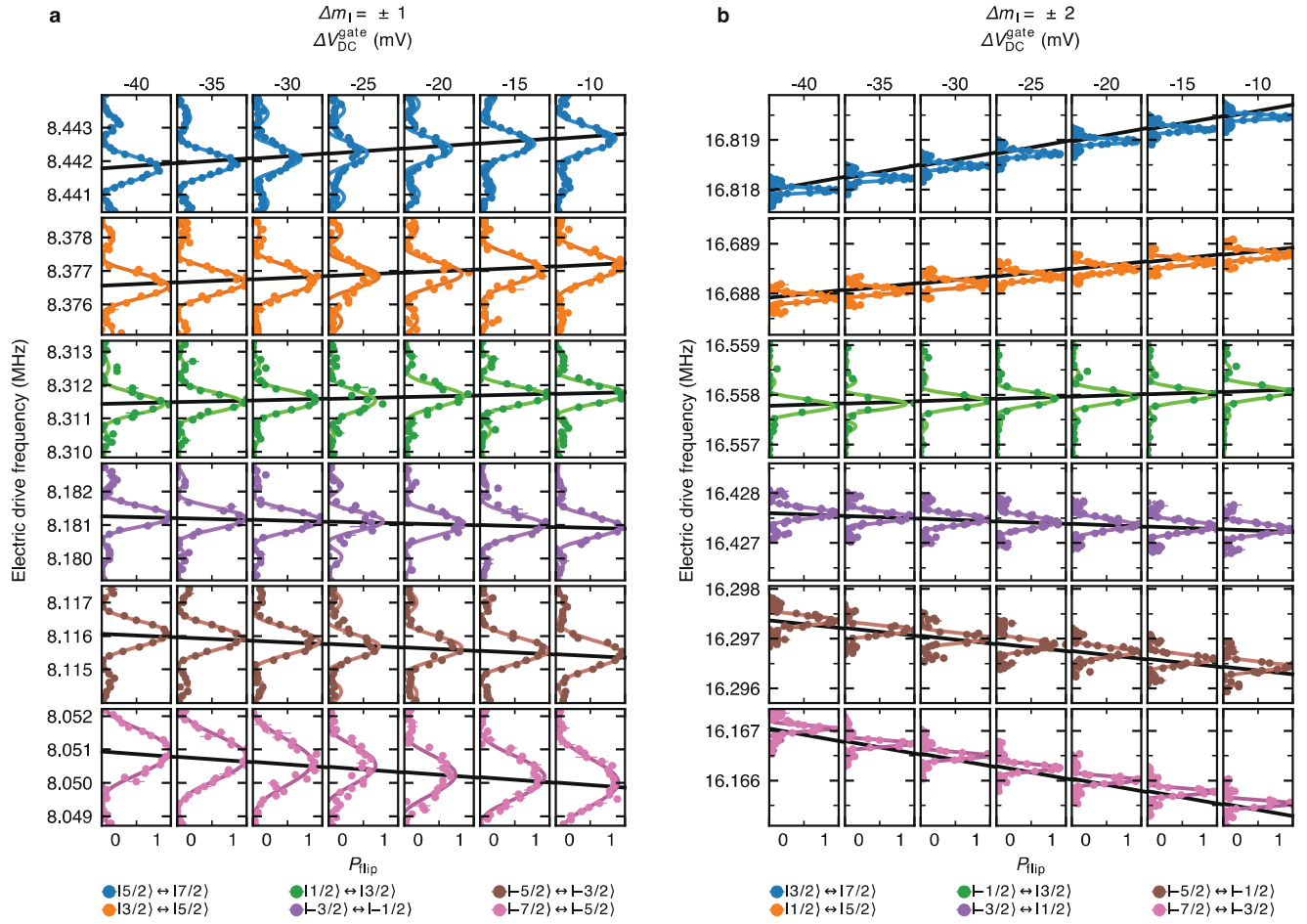
Extended Data Fig. 1 | ESR spectrum in a magnetic field of $B_0 = 1.496$ T. **a, ESR frequencies for eight nuclear states. The average difference between successive ESR transition frequencies (black line) gives a hyperfine interaction of $A = 96.5$ MHz, substantially lower than the bulk value of 101.52 MHz. One**

possible cause for this deviation is strain, which is known to modify the hyperfine interaction³⁹. **b**, ESR spectral lines. For each nuclear state, the nucleus was initialized at the start of each microwave sweep, and adiabatic ESR pulses with 1 MHz frequency deviation were applied to excite the electron.



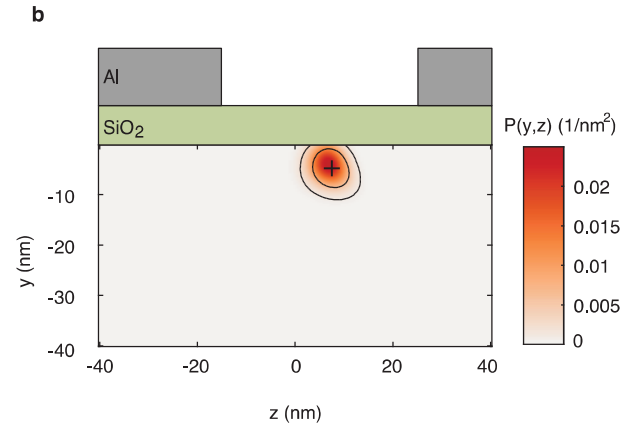
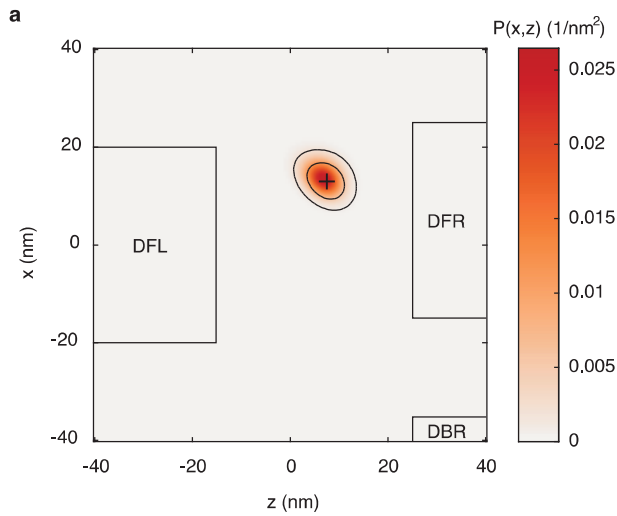
Extended Data Fig. 2 | NER Rabi oscillations on resonance. a, b, Nuclear Rabi oscillations measured with varying NER pulse duration, t_{NER} , while the pulse amplitude was fixed at $V_{\text{gate}}^{\text{RF}} = 20$ mV for $\Delta m_I = \pm 1$ transitions (**a**) and $V_{\text{gate}}^{\text{RF}} = 40$ mV

for $\Delta m_I = \pm 2$ transitions (**b**). Black lines are non-decaying sinusoidal fits to the data, and error bars show the 68% confidence level.



Extended Data Fig. 3 | NER spectral line shifts for varying d.c. gate voltage. **a, b.** The spectral lines of all $\Delta m_l = \pm 1$ transitions (**a**) and $\Delta m_l = \pm 2$ (**b**) transitions are measured while the d.c. gate voltage bias ΔV_{DC}^{gate} is varied (columns) during the NER pulse. We note that this change in V_{DC}^{gate} is applied on top of large gate voltages, of the order of 0.5 V, which are necessary to electrostatically tune the device to enable its operation. The varying ΔV_{DC}^{gate} modifies the quadrupole interaction via the LQSE (see Supplementary Information section 7 for details),

resulting in shifts of the resonance peaks. A single fit to the resonance frequency shifts of all $\Delta m_l = \pm 1$ and $\Delta m_l = \pm 2$ transitions (solid lines) gives an estimate of the gate-dependent quadrupole shift of $\partial f_Q / \partial V_{DC}^{gate} = 9.9(3) \text{ Hz mV}^{-1}$. From the top to the bottom transition, the drive strengths V_{RF}^{gate} are 20 mV, 20 mV, 25 mV, 20 mV and 25 mV for $\Delta m_l = \pm 1$ (**a**) and 30 mV, 30 mV, 40 mV, 40 mV and 40 mV for $\Delta m_l = \pm 2$ (**b**). Error bars show the 68% confidence level.



Extended Data Fig. 4 | Position triangulation of the ^{123}Sb donor. The colour map shows the probability of finding the donor in a certain location. **a, b**, Probability density function found using a least-squares estimation comparing simulated gate-to-donor coupling strengths with the experimentally observed strengths (see Supplementary Information section 7A for details, including the locations of the donor gates DFR, DFL and DBR). To improve on the low resolving power of the triangulation method in the y direction, the triangulation probability density function is multiplied with the donor implantation probability density function (see Supplementary

Information section 7A for details). This has little effect laterally, but greatly confines the likely depth range of the donor within the range expected on the basis of the donor implantation parameters. The most likely donor position, indicated by a cross, is at the lateral location $(x, z) = (13 \text{ nm}, 8 \text{ nm})$ at a depth of $y = -5 \text{ nm}$. The probability density functions are normalized over the model volume and are integrated over the out-of-plane axis in both panels, specifically, $P(x, z) = \int P(\mathbf{r}) dy$ and $P(y, z) = \int P(\mathbf{r}) dx$. The contour lines mark the 68% and 95% confidence regions.

Observation of the Kondo screening cloud

<https://doi.org/10.1038/s41586-020-2058-6>

Received: 20 June 2019

Accepted: 11 December 2019

Published online: 11 March 2020

 Check for updates

Ivan V. Borzenets^{1,6}✉, Jeongmin Shim^{2,6}, Jason C. H. Chen³, Arne Ludwig⁴, Andreas D. Wieck⁴, Seigo Tarucha⁵, H.-S. Sim^{2,6}✉ & Michihisa Yamamoto⁵✉

When a magnetic impurity exists in a metal, conduction electrons form a spin cloud that screens the impurity spin. This basic phenomenon is called the Kondo effect^{1,2}. Unlike electric-charge screening, the spin-screening cloud^{3–6} occurs quantum coherently, forming spin-singlet entanglement with the impurity. Although the spins interact locally around the impurity, the Kondo cloud can theoretically spread out over several micrometres. The cloud has not so far been detected, and so its physical existence—a fundamental aspect of the Kondo effect—remains controversial^{7,8}. Here we present experimental evidence of a Kondo cloud extending over a length of micrometres, comparable to the theoretical length ξ_K . In our device, a Kondo impurity is formed in a quantum dot^{2,9–11}, coupling on one side to a quasi-one-dimensional channel¹² that houses a Fabry–Pérot interferometer of various gate-defined lengths L exceeding one micrometre. When we sweep a voltage on the interferometer end gate—separated by L from the quantum dot—to induce Fabry–Pérot oscillations in conductance we observe oscillations in the measured Kondo temperature T_K , which is a signature of the Kondo cloud at distance L . When L is less than ξ_K the T_K oscillation amplitude becomes larger as L becomes smaller, obeying a scaling function of a single parameter L/ξ_K , whereas when L is greater than ξ_K the oscillation is much weaker. Our results reveal that ξ_K is the only length parameter associated with the Kondo effect, and that the cloud lies mostly within a length of ξ_K . Our experimental method offers a way of detecting the spatial distribution of exotic non-Fermi liquids formed by multiple magnetic impurities or multiple screening channels^{13–16} and of studying spin-correlated systems.

Although Kondo physics for a single magnetic impurity has been well established except for its spatial extension, our understanding of multiple impurity systems such as Kondo lattices, spin glasses, and high-transition-temperature (high- T_c) superconductors is far from complete. In such systems, the Kondo cloud length (or the spatial distribution of a Kondo cloud) with respect to the distance between impurities and other length parameters is crucial for an understanding of their properties. The detection and control of a Kondo cloud is therefore a milestone in condensed matter physics. There have been attempts to detect the Kondo cloud for 50 years^{3–8,12,17–24}. Nuclear magnetic resonance measurements have not found any signature of the cloud⁷. Scanning tunnelling microscopy experiments have shown a signature of the Kondo effect but in a region a distance away from a magnetic impurity that is much shorter than the cloud length¹⁷. The difficulty lies in the fact that measuring spin correlation in the Kondo screening requires the fast detection of tens of gigahertz¹⁸ and there may be complications arising from the atomic or electronic structures of the sample. However, recent advances in nanotechnology have opened up another way of detecting the Kondo cloud. We can now prepare a single spin in a quantum dot (QD) in contact with an electron reservoir, thus achieving systematic control of a single-channel Kondo state^{9,10}. The theoretical value of the cloud length is typically³ $\xi_K = \hbar v_F / (k_B T_K) \approx 1 \mu\text{m}$ (where k_B is the Boltzmann

constant) for $T_K \approx 1 \text{ K}$ and the Fermi velocity is $v_F \approx 10^5 \text{ m s}^{-1}$. A recent theoretical study²⁵ of quantum entanglement shows that the Kondo state lies mostly within the distance ξ_K from the impurity, with a long algebraically decaying tail extending farther. Interesting proposals suggest the use of a finite-size electron reservoir, to observe competition between the cloud length and the reservoir size^{19–21}.

Here, instead of rigidly limiting the reservoir size, we perturb the reservoir by inducing a weak barrier at a position L far from the Kondo impurity and observe the resulting change in the Kondo effect with varying barrier position, following a recent proposal¹². Figure 1a shows the device and measurement schemes. An unpaired electron spin (magnetic impurity) is confined in a QD coupled to a one-dimensional (1D), long, ballistic channel^{26,27}. The QD is in the Coulomb blockade regime. The 1D channel is tuned to contain several conducting channels and has three quantum point contact (QPC) gates placed away from the QD at lengths $L = 1.4 \mu\text{m}$, $3.6 \mu\text{m}$ and $6.1 \mu\text{m}$. Application of voltage V_{QPC} to one of the QPC gates creates a weak barrier so that a Fabry–Pérot (FP) cavity²⁸ of length L is formed between the QD and the QPC. The charging energy of the FP cavity is ineffective owing to strong coupling to the reservoir through multiple conducting channels. Changing V_{QPC} allows us to continuously tune the FP cavity between on- and off-resonances by altering the cavity length on the scale of the Fermi wavelength $\Delta L \approx \lambda_F$ (where

¹Department of Physics, City University of Hong Kong, Kowloon, Hong Kong. ²Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea.

³Department of Applied Physics, University of Tokyo, Tokyo, Japan. ⁴Chair of Applied Solid State Physics, Faculty of Physics and Astronomy, Ruhr-University Bochum, Bochum, Germany.

⁵Center for Emergent Matter Science (CEMS), RIKEN, Saitama, Japan. ⁶These authors contributed equally: I. V. Borzenets, J. Shim. ✉e-mail: iborzenet@cityu.edu.hk; hssim@kaist.ac.kr;

michihisa.yamamoto@riken.jp

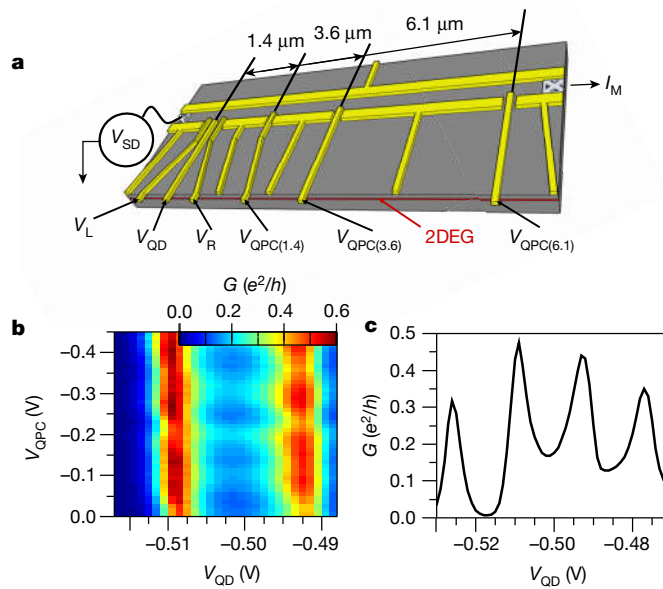


Fig. 1 | Measurement setup and characterization. **a**, Device and measurement schematic. The device consists of a QD coupled to a 1D channel (see Methods), in which three QPC gates are embedded at distances $L = 1.4 \mu\text{m}$, $3.6 \mu\text{m}$ and $6.1 \mu\text{m}$ from the QD. The activation of a QPC gate creates a FP cavity of length L . The QD is tuned via a central plunger gate voltage V_{QD} and two side gates V_L and V_R . The device is measured via the lock-in method: a small a.c. voltage V_{SD} is applied, and the current I_M through the system is measured. **b**, The conductance G (measured in units of quantum conductance, e^2/h) of the device versus the plunger gate voltage V_{QD} and the $L = 1.4 \mu\text{m}$ QPC gate voltage V_{QPC} . Coulomb blockade peaks are observed with respect to changing V_{QD} . Oscillations associated with the FP cavity are seen with respect to changing V_{QPC} . **c**, Conductance G versus V_{QD} taken at $V_{\text{QPC}} = 0$. Coulomb blockade and a region of enhanced conductance around $V_{\text{QD}} = -0.50 \text{ V}$ associated with the Kondo valley are clearly observed.

$\lambda_F \approx 40 \text{ nm} \ll L, \xi_K$) without affecting the potential profile around the QD. We find that changes in V_{QPC} strongly affect the measured Kondo temperature T_K when the cavity length L is shorter than $\xi_{K\infty} = \hbar v_F / (k_B T_{K\infty})$. Here $\xi_{K\infty}$ and $T_{K\infty}$ are the bare theoretical cloud length and bare Kondo temperature defined in the absence of the QPCs or equivalently for the case of $L = \infty$. For $L \gg \xi_{K\infty}$, changes of V_{QPC} have little effect on T_K . This implies that the Kondo state extends over about $\xi_{K\infty}$.

The device is defined using top gates deposited on top of a GaAs/AlGaAs two-dimensional electron gas (2DEG) wafer, with an electron mean free path of about $8 \mu\text{m}$ (which is bigger than the device size). The QD population is controlled via a middle plunger gate voltage V_{QD} . The coupling of the QD to the 1D channel on the right side and to the left side is adjusted by changing the side gate voltages V_R and V_L , respectively. The QD is coupled more strongly to the right 1D channel than the left channel, so that the Kondo state is sensitive to the FP cavity of the 1D channel. We tune V_R to change the bare cloud length $\xi_{K\infty}$ and Kondo temperature $T_{K\infty}$. Figure 1c shows the conductance G between the right 1D channel and the left lead via the QD as a function of the plunger gate voltage V_{QD} when the QPCs are turned off. Several Coulomb blockade peaks are clearly visible with the measured charging energy of $>500 \mu\text{eV}$ (ref. ²⁹). The Kondo effect is observed, manifesting itself in increased conductance in the valley region between two Coulomb blockade peaks^{2,9,10}. The effect of the QPC gates is shown in Fig. 1b, which plots the conductance G versus plunger gate voltage V_{QD} as well as the voltage V_{QPC} applied to the QPC gate at $L = 1.4 \mu\text{m}$. Both the Coulomb blockade peaks as well as the Kondo valley undergo FP oscillations with respect to changing V_{QPC} ; we note that at small V_{QPC} (equivalent to the first few FP oscillations) there is no effect on the QD energy level. The resonance level spacing $\Delta \approx 300 \mu\text{eV}$ estimated

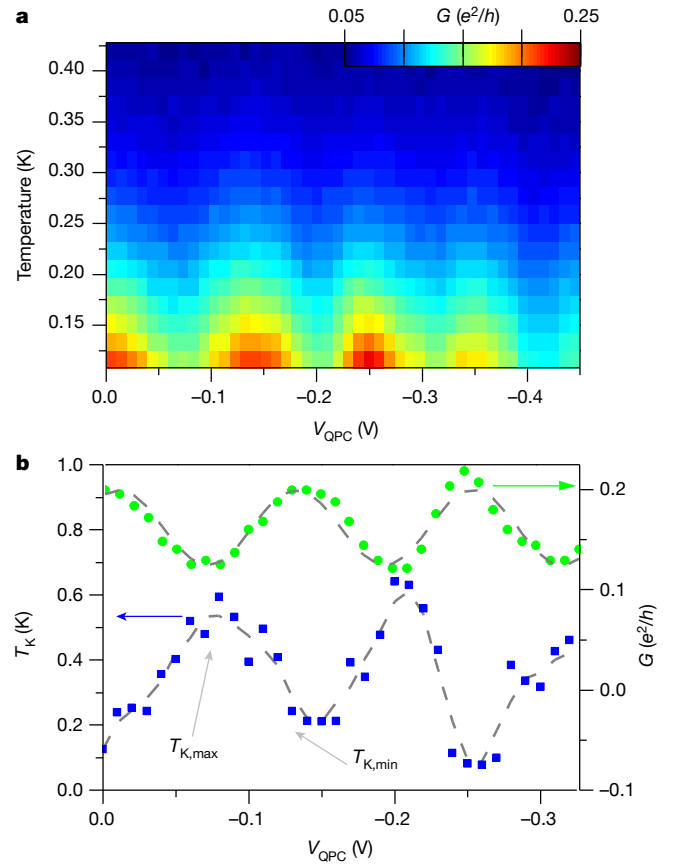


Fig. 2 | Influence of FP interference on the Kondo effect. **a**, Conductance G at the centre of the Kondo valley versus the QPC gate voltage V_{QPC} at $L = 1.4 \mu\text{m}$ and device temperature T . G decreases with increasing T , indicative of the Kondo effect. For each V_{QPC} , the Kondo temperature T_K is extracted by fitting conductance versus temperature to an empirical formula³⁰. **b**, Plot of the extracted T_K versus V_{QPC} (blue), shown alongside the conductance of the Kondo valley centre at base temperature G (green). T_K oscillates with respect to V_{QPC} , but in anti-phase with respect to conductance. The T_K oscillation amplitude is quantified by tracking the maximum $T_{K,\text{max}}$ and minimum $T_{K,\text{min}}$ of the first oscillation.

from the data is consistent with the cavity length $L = 1.4 \mu\text{m}$, considering that $v_F = 2.46 \times 10^5 \text{ m s}^{-1}$. The resonance level broadening of about $80 \mu\text{eV}$ implies a weak barrier formed by V_{QPC} (Supplementary Information).

To see the effect of the FP oscillations on the Kondo state, we measure the conductance G around the centre of the Kondo valley as a function of the QPC gate voltage V_{QPC} at different temperatures T (Fig. 2a). The valley conductance decreases with increasing temperature, indicative of the Kondo state. For each value of V_{QPC} we extract T_K around the valley centre by fitting conductance G versus temperature T to a well known empirical formula (see Methods)³⁰. This method of extracting T_K is applicable for constant density of states of the reservoirs, but is still applicable to our QD coupled to the FP cavity, because the resonance level broadening is sufficiently large or the QPC barrier is weak (see Methods). We find that the Kondo temperature T_K undergoes oscillations with respect to changing V_{QPC} (Fig. 2b). Clearly, the Kondo state is affected by the perturbation at a location micrometres distant from the QD. The oscillation shows that the electron density of the FP cavity at the Fermi level that is coupled with the Kondo impurity differs between on and off resonances. This implies that the Kondo coherence is extended through the entire FP cavity, supporting the picture of the spatial extension of the Kondo cloud. It also implies that the resonance level spacing Δ is larger than the Kondo temperature, as theoretically expected. We note that, based on the scatter in T_K versus

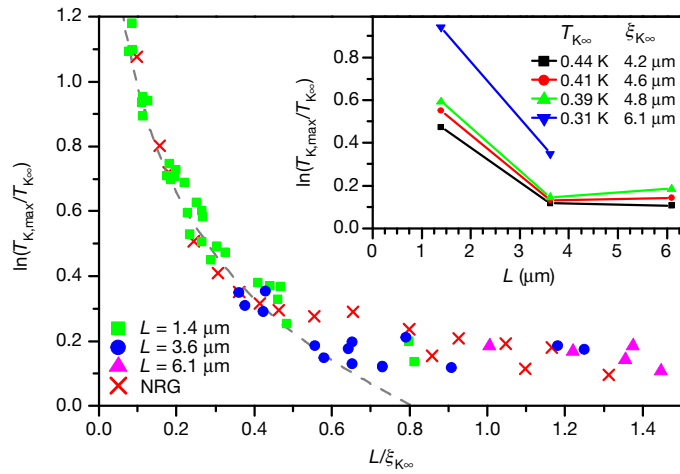


Fig. 3 | Shape of the Kondo cloud revealed from modulation of the Kondo temperature. Oscillation amplitude of T_K as a function of the FP cavity length L . The amplitude is quantified by $\ln(T_{K,\max}/T_{K,\infty})$, the maximum value $T_{K,\max}$ of the oscillation normalized by the bare Kondo temperature $T_{K,\infty}$ in logarithmic scale, and L is scaled by the bare Kondo cloud length $\xi_{K,\infty}$. We note that $T_{K,\infty}$ is estimated as $T_{K,\infty} = (T_{K,\max}T_{K,\min})^{1/2}$ (see Supplementary Information). The green squares are obtained from the oscillation of T_K with respect to changes in the voltage V_{QPC} of the QPC gate located at $L = 1.4 \mu\text{m}$ from the QD, the blue circles are for $L = 3.6 \mu\text{m}$, and the magenta triangles are for $L = 6.1 \mu\text{m}$. Different data points of the same symbol correspond to different $T_{K,\infty}$ (that is, different settings of V_R and V_L). The data are compared with the theoretical results of the NRG calculation (red crosses) and the scaling of $\ln(T_{K,\max}/T_{K,\infty}) = -\eta \ln(L/\xi_{K,\infty})$ with $\eta = 0.47$ (dashed curve). The inset shows the same data as the oscillation amplitude of T_K versus the length L not scaled by $T_{K,\infty}$. The different curves represent those having the same $T_{K,\infty}$.

V_{QPC} (where a smooth curve is expected), we estimate an error for the extracted T_K of about 20%.

The oscillations of T_K are anti-phase with those of the conductance. This agrees with scattering theories combined with the Fermi liquid and numerical renormalization group (NRG) methods when the QD is coupled more strongly to the FP cavity than to the left channel (Supplementary Information). In the on-resonance situations where the Fermi momentum k_F satisfies $e^{i2k_F L} = 1$, the FP cavity supports a maximum amount of electrons at the Fermi level, and hence the Kondo state with maximum Kondo temperature is developed. The development effectively changes the resonance condition to off-resonance, resulting in a minimum value of the conductance, since the electrons gain twice the scattering phase shift $\pi/2$ off the Kondo impurity in the QD³¹. The opposite happens in the off-resonance situations of $e^{i2k_F L} = -1$. We note that evidence of the Kondo scattering phase $\pi/2$ has been observed²⁶, and that the antiphase in Fig. 2b can be considered as further evidence. To quantify the effect of the FP cavity on the Kondo state we track the maximum $T_{K,\max}$ and minimum $T_{K,\min}$ of the first oscillation of T_K . Only the first oscillation is used because the subsequent oscillations require a higher V_{QPC} , which will have a stronger effect on the QD owing to capacitive coupling through the FP interferometer island.

We now discuss the dependence of the oscillation amplitude of T_K on the FP cavity length L shown in the inset of Fig. 3. The amplitude is quantified by $\ln(T_{K,\max}/T_{K,\infty})$, and we estimate the bare Kondo temperature from the oscillation as $T_{K,\infty} = (T_{K,\max}T_{K,\min})^{1/2}$ (see Supplementary Information for a validation of the estimate); $T_{K,\infty}$ is not directly accessible because a QPC gate can form a barrier even when its voltage is turned off. The oscillation amplitude is small, that is, $T_{K,\max} \approx 1.1T_{K,\infty}$ or $T_{K,\max} \approx 1.2T_{K,\min}$ for cavity lengths $L = 3.6 \mu\text{m}$ and $6.1 \mu\text{m}$ and $T_{K,\infty} = 0.39$ – 0.44 K. By contrast, the amplitude becomes much larger for $L = 1.4 \mu\text{m}$, for example, $T_{K,\max} \approx 3.2T_{K,\infty}$ or $T_{K,\max} \approx 10T_{K,\min}$ for $T_{K,\infty} = 0.31$ K. For each

L , the amplitude becomes larger as $T_{K,\infty}$ is smaller. The result shows that the Kondo state is sensitive (oscillation amplitude of $>30\%$) to the perturbation at distance $L \lesssim 3.6 \mu\text{m}$ for $T_{K,\infty} = 0.39$ – 0.44 K, while also sensitive at $L > 3.6 \mu\text{m}$ for $T_{K,\infty} = 0.31$ K. This implies that the cloud length is close to $3.6 \mu\text{m}$ for $T_{K,\infty} = 0.39$ – 0.44 K, while larger than $3.6 \mu\text{m}$ for $T_{K,\infty} = 0.31$ K. This finding is consistent with the bare cloud length $\xi_{K,\infty}$ estimated by the theoretical relation of $\xi_{K,\infty} = \hbar v_F / (k_B T_{K,\infty})$, that is, $\xi_{K,\infty} = 5.19 \mu\text{m}$, $4.12 \mu\text{m}$, $3.92 \mu\text{m}$ and $3.65 \mu\text{m}$ for $T_{K,\infty} = 0.31$ K, 0.39 K, 0.41 K and 0.44 K, respectively.

To see the universality of the results, we plot the oscillation amplitude of T_K versus the cavity length L scaled by the bare cloud length $\xi_{K,\infty}$ in Fig. 3. Since we analyse only the first oscillation, the transmission through the QPC is almost independent of V_{QPC} and common for all QPCs. We find that all data points fall onto a single curve, as theoretically expected for a fixed transmission through the QPC. This scaling result is the evidence that ξ_K is the only length parameter associated with the Kondo effect. For $L \geq \xi_{K,\infty}$, the Kondo state is little affected by the perturbation at distance L , because the maximum of the oscillation is 20% larger than the minimum, $T_{K,\max} \approx 1.2T_{K,\min}$. As L decreases below $\xi_{K,\infty}$, the oscillation amplitude becomes very much larger, showing that the maximum is 1,000% larger than the minimum ($T_{K,\max} \approx 10T_{K,\min}$) at $L \approx 0.1\xi_{K,\infty}$ (we note that the oscillation amplitude is much bigger than the 20% error in the extracted T_K). The increase follows the universal scaling¹² of $\ln(T_{K,\max}/T_{K,\infty}) = -\eta \ln(L/\xi_{K,\infty})$ with a constant η defined as the modulation of the density of states set by the QPC pinch-off strength. The plot is in good agreement with theoretical NRG calculations based on realistic parameters estimated from sample characterization (Supplementary Information). The result is consistent with the theoretical result²⁵ of the spatial distribution of the Kondo singlet entanglement that the main body of the Kondo cloud lies inside the length $\xi_{K,\infty}$ with a long tail extending beyond $\xi_{K,\infty}$. An equivalent, alternative picture is that for a FP cavity of length L , there are $L/\xi_{K,\infty}$ (approximately $k_B T_{K,\infty} / \Delta$) localized single-particle states of size $\xi_{K,\infty}$ in a row. When $L/\xi_{K,\infty} > 1$, the single-particle state located closest to the Kondo impurity forms the main body of the cloud and is wholly within L . Hence, the local perturbation at L affects only the other single-particle states contributing to the cloud tail. When $L/\xi_{K,\infty} < 1$, the main body of the cloud extends beyond distance L and hence it is strongly affected by the perturbation.

Our result provides evidence of the spatial distribution of the Kondo state over micrometres. It will be interesting to study the spatial distribution further, for example, by engineering the spatial spin screening and the entanglement of the Kondo state. For example, by applying large QPC gate voltage to our device, we could systematically study the screening cloud of a Kondo box^{19–21}. It will also be valuable to study the spin screening by multiple independent channels as in the multi-channel Kondo effects or in a situation of multiple impurities as in the two-impurity Kondo effects, because those effects are accompanied by non-Fermi liquids and a quantum phase transition^{13,14}. Our strategy of detecting spin screening by applying a weak electrostatic gate at a position distant from an impurity spin is applicable to the realization^{15,16} of these effects with systematic control.

Our work is an initial measurement of a Kondo cloud. It may enable Kondo cloud detection in other systems and in other ways, such as by measuring spin–spin correlations. The universality implies that Kondo clouds in a conventional metal or in the mixed-valence regime have a spatial distribution similar to that found here, albeit with a shorter ξ_K . Ballistic electron transport over ξ_K will be required to apply our approach.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

1. Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1993).
2. Glazman, L. I. & Pustilnik, M. Low-temperature transport through a quantum dot. In *Nanophysics: Coherence and Transport* (ed. Bouchiat, H. et al.) 427–478 (Elsevier, 2005).
3. Affleck, I. The Kondo screening cloud: what it is and how to observe it. In *Perspectives of Mesoscopic Physics* (eds Aharony, A. & Entin-Wohlman, O.) Ch. 1, 1–44 (World Scientific Publishing, 2010).
4. Grüner, G. & Zawadowski, A. Magnetic impurities in non-magnetic metals. *Rep. Prog. Phys.* **37**, 1497–1583 (1974).
5. Gubernatis, J. E., Hirsch, J. E. & Scalapino, D. J. Spin and charge correlations around an Anderson magnetic impurity. *Phys. Rev. B* **35**, 8478 (1987).
6. Barzykin, V. & Affleck, I. The Kondo screening cloud: what can we learn from perturbation theory? *Phys. Rev. Lett.* **76**, 4959 (1996).
7. Boyce, J. P. & Slichter, C. P. Conduction-electron spin density around Fe impurities in Cu above and below T_K . *Phys. Rev. Lett.* **32**, 61 (1974).
8. Sørensen, E. S. & Affleck, I. Scaling theory of the Kondo screening cloud. *Phys. Rev. B* **53**, 9153 (1996).
9. Goldhaber-Gordon, D., Shtrikman, H., Mahalu, D., Abusch-Magder, D. & Meirav, U. Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
10. Cronenwett, S. M., Oosterkamp, T. H. & Kouwenhoven, L. P. A tunable Kondo effect in quantum dots. *Science* **281**, 540–544 (1998).
11. Kouwenhoven, L. P. & Glazman, L. I. Revival of the Kondo effect. *Phys. World* **14**, 33–38 (2001).
12. Park, J., Lee, S.-S. B., Oreg, Y. & Sim, H.-S. How to directly measure a Kondo cloud's length. *Phys. Rev. Lett.* **110**, 246603 (2013).
13. Cox, D. L. & Jarrell, M. The two-channel Kondo route to non-Fermi-liquid metals. *J. Phys. Condens. Matter* **8**, 9825–9853 (1996).
14. Affleck, I. Non-Fermi liquid behavior in Kondo models. *J. Phys. Soc. Jpn* **74**, 59–66 (2005).
15. Potok, R. M., Rau, I. G., Shtrikman, H., Oreg, Y. & Goldhaber-Gordon, D. Observation of the two-channel Kondo effect. *Nature* **446**, 167–171 (2007).
16. Iftikhar, Z. et al. Two-channel Kondo effect and renormalization flow with macroscopic quantum charge states. *Nature* **526**, 233–236 (2015).
17. Prüser, H. et al. Long-range Kondo signature of a single magnetic impurity. *Nat. Phys.* **7**, 203–206 (2011).
18. Borda, L. Kondo screening cloud in a one-dimensional wire: numerical renormalization group study. *Phys. Rev. B* **75**, 041307 (2007).
19. Thimm, W. B., Kroha, J. & von Delft, J. Kondo box: a magnetic impurity in an ultrasmall metallic grain. *Phys. Rev. Lett.* **82**, 2143 (1999).
20. Simon, P. & Affleck, I. Finite-size effects in conductance measurements on quantum dots. *Phys. Rev. Lett.* **89**, 206602 (2002).
21. Bomze, Yu. et al. Two-stage Kondo effect and Kondo-box level spectroscopy in a carbon nanotube. *Phys. Rev. B* **82**, 161411 (2010).
22. Holzner, A., McCulloch, I. P., Schollwöck, U., von Delft, J. & Heidrich-Meisner, F. Kondo screening cloud in the single-impurity Anderson model: a density matrix renormalization group study. *Phys. Rev. B* **80**, 205114 (2009).
23. Büsser, C. A. et al. Numerical analysis of the spatial range of the Kondo effect. *Phys. Rev. B* **81**, 045111 (2010).
24. Mitchell, A. K., Becker, M. & Bulla, R. Real-space renormalization group flow in quantum impurity systems: local moment formation and the Kondo screening cloud. *Phys. Rev. B* **84**, 115120 (2011).
25. Lee, S.-S. B., Park, J. & Sim, H.-S. Macroscopic quantum entanglement of a Kondo cloud at finite temperature. *Phys. Rev. Lett.* **114**, 057203 (2015).
26. Takada, S. et al. Transmission phase in the Kondo regime revealed in a two-path interferometer. *Phys. Rev. Lett.* **113**, 126601 (2014).
27. van der Wiel, W. G. et al. The Kondo effect in the unitary limit. *Science* **289**, 2105–2108 (2000).
28. van Houten, H. et al. Coherent electron focusing with quantum point contacts in a two-dimensional electron gas. *Phys. Rev. B* **39**, 8556 (1989).
29. Kouwenhoven, L. P. et al. Electron transport in quantum dots. In *Mesoscopic Electron Transport* (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) NATO Advanced Study Institutes Series E, Vol. 345, 105–214 (Kluwer Academic, 1997).
30. Goldhaber-Gordon, D. et al. From the Kondo regime to the mixed-valence regime in a single-electron transistor. *Phys. Rev. Lett.* **81**, 5225 (1998).
31. Nozières, P. A Fermi-liquid description of the Kondo problem at low temperatures. *J. Low Temp. Phys.* **17**, 31–42 (1974).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Sample preparation

The system of the QD coupled to the quasi-1D FP resonant cavity was fabricated in a 2DEG (carrier density $n = 3.12 \times 10^{11} \text{ cm}^{-2}$, mobility $\mu = 0.86 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) heterointerface using a standard surface Schottky gate technique.

Measurement

Measurements were performed in an Oxford Instruments MX100 dilution refrigerator with a base lattice temperature of 40 mK. The base electron temperature was measured to be around 80 mK. (Electron temperature versus measured mixing chamber temperature was calibrated via analysis of the QD Coulomb blockade peak width.) The quasi-1D channels were formed by applying an approximately -0.45 V (-1 V) signal to the gates defining the bottom (top) boundary of the channel. (This voltage was sufficient to fully deplete the carriers below the gates, isolating the channel from the rest of the 2DEG.) Electron transport was measured using the lock-in method. An a.c. voltage oscillation with a d.c. offset was applied to a sample via a divider with the a.c. excitation set to $3\text{--}15 \text{ } \mu\text{V}$. The d.c. offset was set so as to achieve zero bias at the sample. (The d.c. offset was varied in order to calculate the QD charging energy.) Current through the sample was measured using a custom current sense amplifier with a current-sense resistor of $10 \text{ k}\Omega$ mounted on the mixing chamber. Gate voltages were controlled via a custom digital-to-analog converter (DAC). Temperature of the device was changed globally via a heating coil at the mixing chamber.

We provide additional details about the gate voltage settings (see Fig. 1a). The voltage applied to the long global gate defining the upper boundary of the 1D channels and the QD was fixed to about -1.0 V for the entire experiment. The gate voltages defining the left and right 1D channels from the lower side were set to about -0.45 V , fully depleting the carriers underneath the gates but leaving several conducting channels in each 1D wire. The gate voltages defining the QD were operated around -0.5 V , being varied in order to achieve different values of T_{K} . The voltages on the tunnel coupling gates $V_{\text{L,R}}$ were set in a range of -0.45 V to -0.6 V , while the middle plunger gate voltage V_{QD} was swept between -0.4 V and -0.7 V . The QPC gates were set to zero voltage when not operated. We confirmed that $V_{\text{QPC}(1,4)}$ and $V_{\text{QPC}(3,6)}$ have no influence on the conductance through longer FP cavities defined by $V_{\text{QPC}(3,6)}$ or $V_{\text{QPC}(6,1)}$.

Calibration of measured sample temperature

For temperatures below about 500 mK it becomes increasingly difficult to thermalize hot electrons sent to the sample from the measurement electronics. In this measurement setup, the hot electrons were cooled at the mixing chamber via a thermalization coil as well as via copper powder filters. Nevertheless, calibration was required to match the measured mixing chamber lattice temperature to the actual electron temperature in the sample. This was achieved by measuring the shape of the Coulomb blockade peaks of the QD with all the QPCs turned off (Extended Data Fig. 1c). Picking a peak that does not feature any Kondo temperature, we should expect that the conductance $G(\delta V_{\text{G}})$ (where δV_{G} is the shift in plunger gate voltage away from the Coulomb blockade peak centre) will be proportional to $\cosh^{-2}\left(\frac{a\delta V_{\text{G}}}{2k_{\text{B}}T}\right)$ (ref. 29). The constant a is independent of electron temperature and should depend only on the sample geometry. (Extended Data Fig. 1a shows the fit of G versus gate voltage V_{G} taken at a measured temperature of $T_{\text{measured}} = 300 \text{ mK}$). Indeed, above $T_{\text{measured}} > 600 \text{ mK}$ we see that a has little to no variation, so the mixing chamber thermometer temperature and the electron temperature are the same $T_{\text{calibrated}} = T_{\text{measured}}$. By looking at how the constant a evolves for $T_{\text{measured}} < 600 \text{ mK}$ we can now build the calibration for the electron temperature $T_{\text{calibrated}} = F(T_{\text{measured}})$ as shown in Extended Data Fig. 1b. Throughout the text, the temperature T refers to the calibrated temperature $T_{\text{calibrated}}$.

QPC barrier strengths

The strengths of the oscillations in $T_{\text{K,max}}/T_{\text{K,min}}$ depends on the QPC pinch-off strength α . Here we define $\alpha = 1 - (t_0/t)^2$ with t_0 being the hopping energy across the QPC and t being the hopping energy along the 1D channel. ($\alpha = 0$ means there is no pinchoff present due to the QPC, and $\alpha = 1$ means that the QPC fully decouples the FP cavity from the rest of the wire. See the Hamiltonian in the Supplementary Information.) Below we estimate α in three different ways, which show $\alpha \approx 0.1$. Fitting the main results (Fig. 3) to the NRG calculations, we arrive at $\alpha \approx 0.1$. In addition, we find that α can be obtained from the ratio of on- and off-resonance conductances of the Kondo valley at $T = 0$ as follows: $\alpha = 1 - (G_{0,\text{Min}}/G_{0,\text{Max}})^{0.5}$. The conductances at zero temperature can be extracted similarly to T_{K} by fitting to the empirical formula³⁰ (see below for fitting details). Averaging the data over several values of V_{L} and V_{R} we arrive at $\alpha \approx 0.1$. Finally, we estimate α also by considering the behaviour of the system away from the Kondo regime. The strength is proportional to the fluctuations in the local carrier density ρ inside the FP cavity: $\alpha = 1 - (\rho_{\text{min}}/\rho_{\text{max}})^{0.5}$, where $\rho_{\text{max(min)}}$ is the maximum (minimum) value of the fluctuations in ρ . In turn, the effective coupling strength Γ_{r} of the QD to the FP cavity is proportional to the local carrier density ρ . The coupling strength Γ_{r} directly affects the width of the Coulomb blockade peak with respect to gate voltage V_{G} . Indeed, as we tune the FP resonance by changing the QPC gate voltage V_{QPC} , the Coulomb blockade peak undergoes fluctuations in width in synchrony with fluctuations in conductance (Extended Data Fig. 2). Taking the first oscillation, we arrive at the QPC pinchoff strength $\alpha = 1 - (W_{\text{min}}/W_{\text{max}})^{0.5} \approx 0.1$, where W is the width of the oscillation.

For the case of ballistic electron transport and a similar α for all three QPC gates, one would expect the amplitude of the FP oscillations in Coulomb-blockade peak conductance to be independent of the QPC gates at different distances. Indeed, when comparing the oscillations for the same settings of V_{L} and V_{R} , but activating QPC gates at the three different distances, we see that the amplitudes are within 15% of each other (Extended Data Table 1).

Kondo temperature estimation

The Kondo temperature T_{K} is estimated by fitting the experimental data to the empirical formula³⁰ of

$$G(T) = G_0 \left(\frac{T_{\text{K}}^2}{T^2 + T_{\text{K}}^2} \right)^s$$

with $T'_{\text{K}} = T_{\text{K}}/\sqrt{2^{1/s} - 1}$, where the zero-temperature conductance G_0 and the exponent s are fitting parameters; according to ref. 30, the exponent s should not vary much from 0.22; in this sense, the fitting has only two fitting parameters, T_{K} and G_0 . This estimation works well when the density of states of the reservoirs coupled to a Kondo impurity is approximately energy-independent near the Fermi level. In our setup, in which the Kondo impurity in the QD is coupled to the FP cavity, the density of states is energy-dependent and the estimation is applicable when the QPC barrier defining the FP cavity is so weak that the energy dependence is not crucial. The applicability of the estimation is confirmed by our NRG calculation with the model parameters chosen from the experimental data (Supplementary Information).

An example of the fit of our data to the empirical formula is shown in Extended Data Fig. 3, with T_{K} and G_0 as fitting parameters (Extended Data Table 2). (On average, the fitting function returned a degrees-of-freedom adjusted R^2 value of 0.995 for all the values of V_{QPC} within the first oscillation cycle.) The trend of G versus T is described well by the empirical formula with $s = 0.22 \pm 0.01$ (Extended Data Table 2). The fit is done with the temperature dependence of the conductance G over an electron temperature window of $0.1\text{--}0.5 \text{ K}$. In usual cases (namely, in the absence of the FP cavity), the conductance G measured within the temperature window $[0.5T_{\text{K}}, 1.5T_{\text{K}}]$ around the T_{K} is sufficient to

correctly estimate the Kondo temperature from the fit; this can be confirmed by using NRG calculations or by analysing the empirical formula. It is because this window can capture the Kondo crossover (from the Kondo fixed point to the local moment fixed point). Even with the FP cavity, this estimation is expected to work well in the regime of $L < 0.4\xi_{K\infty}$ according to our NRG calculation (Supplementary Information). The estimation becomes worse as L increases ($>0.4\xi_{K\infty}$), but not enough to affect our conclusion in the main text and Fig. 3.

We remark two points. First, while the fitting function implies a good fit, it is difficult to estimate the error in the fitted T_K from a single fit. Instead, as discussed in the main text, we obtain an error by looking at the scatter of T_K when the QPC voltage is changed. In an ideal case, the Kondo temperature is expected to evolve smoothly with QPC voltage. That is, for a small change in voltage ΔV_{QPC} we expect only a small deviation in temperature $T_K(V_{QPC}) \approx T_K(V_{QPC} + \Delta V_{QPC})$. Large changes in T_K for small changes in V_{QPC} are therefore attributed to error. Using this method, we arrive at an error in T_K of about 20%.

Second, the fitting parameter $s = 0.22$ implies that our experiment is in the Kondo regime rather than in the mixed-valence regime. The empirical formula provides a good fit with its parameter $s = 0.22$ in the Kondo regime, while the parameter deviates rapidly from 0.22 in the mixed-valence regime³⁰.

We note that the estimation of T_K directly from the experimental data is a merit of our experimental regime; it is unclear how to estimate T_K directly in the opposite regime where the QPC barrier is strong (corresponding to a finite-size reservoir coupled to a QD^{19–21}). This is because the temperature dependence of the conductance G has a nontrivial feature due to the FP resonance when the QPC barrier is strong (the nontrivial feature is enhanced as the barrier becomes stronger, and the feature is not captured by the empirical formula), as shown in our NRG calculations.

Modelling

To model our experiments, we theoretically study an Anderson impurity formed in a QD coupled to two 1D leads, one of which houses a FP cavity. The Hamiltonian of the model is found in the Supplementary Information. The parameters of the model are estimated from the experimental data. The QD spectral function is obtained by using the NRG method, and the temperature dependence of the conductance is computed¹ by combining the spectral function and the Fermi–Dirac distribution function. The feature in Fig. 2 (that the oscillations of T_K are anti-phase with those of the conductance) is reproduced in our theory in two different ways. The feature is obtained by using the scattering-matrix formalism combined with the Fermi liquid theory and by taking into account the scattering phase shift $\pi/2$ off the Kondo impurity. The anti-phase feature is also found in an independent way based on the NRG method and conductance calculation. On the other hand, the NRG result of the Kondo temperature T_K in Fig. 3 is obtained, in the same way as for the experimental estimation of T_K , by fitting the conductance obtained by the NRG method to the empirical formula. The universal scaling behaviour in Fig. 3 is obtained by using the Poor Man’s Scaling¹² and it is confirmed by NRG calculations; the Poor Man’s Scaling and the NRG give the same scaling form but with different values of the parameter

η , as η depends on the estimation method of T_K . The details are found in the Supplementary Information.

Universal scaling at $L < \xi_{K\infty}$

The experimental data in Fig. 3 show the behaviour of

$$\ln(T_{K,\max}/T_{K\infty}) \approx -\eta \ln(L/\xi_{K\infty}) \quad (1)$$

at $L < \xi_{K\infty}$. This behaviour is scaled only by $T_{K\infty}$ and $\xi_{K\infty}$, so it is a universal feature characterizing the core region ($\hbar v_F/U < L < \xi_{K\infty}$) of the Kondo cloud. Here U is the QD charging energy (see Supplementary Information).

We derive the behaviour. In our experimental regime of $t_1 \ll t_i$ and small α , the Poor Man’s Scaling¹² leads to $\ln(T_{K,\max}/T_{K\infty}) \approx -\alpha \ln(L/\xi_{K\infty})$. Here t_i (t_i) is the tunnelling amplitude between the QD and the left channel (the FP cavity). See Supplementary Information. In this estimation based on the Poor Man’s Scaling, η equals α . This implies that the coefficient η is determined mainly by the resonance broadening parameter α for small α . The NRG calculation confirms the universal scaling in equation. (1), but with $\eta \neq \alpha$. The value of η depends on the estimation method of the Kondo temperature; the Kondo temperature is defined in the Poor Man’s Scaling in a way that is different from that in the NRG method. We find that the scaling behaviour also occurs in the other regime of $t_{1,r}$ and α including the regimes of large α . We also note that at $L > \xi_{K\infty}$, the Kondo cloud has a long tail following another algebraic scaling law characterized by quantum entanglement or electron conductance^{25,32}.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

32. Yoo, G., Lee, S.-S. B. & Sim, H.-S. Detecting Kondo entanglement by electron conductance. *Phys. Rev. Lett.* **120**, 146801 (2018).

Acknowledgements I.V.B. acknowledges CityU New Research Initiatives/Infrastructure Support from Central (APRC) (grant number 9610395), and the Hong Kong Research Grants Council (ECS) Project (grant number 21301818). S.T. and M.Y. acknowledge KAKENHI (grant number 38000131). M.Y. acknowledges KAKENHI (grant number 18H04284) and CREST-JST (grant number JPMJCR1876). S.T. acknowledges CREST-JST (grant number JPMJCR1675). M.Y. acknowledges discussions with R. Sakano. H.-S.S. acknowledges support by Korea NRF via the SRC Center for Quantum Coherence in Condensed Matter (grant number 2016R1A5A1008184). A.L. and A.D.W. acknowledge support from DFG-TRR160, BMBF—Q.Link.X16KIS0867 and DFH/UFA CDF-05-06.

Author contributions I.V.B. performed the experimental measurements and analysed experimental data. J.S. and H.-S. S. performed the theoretical calculations. J.C.H.C. fabricated and characterized the device. A.L. and A.D.W. designed and grew the 2DEG wafer. M.Y. designed the experiment. M.Y. and S.T. supervised the project. All authors were involved in discussing results and preparing the manuscript.

Competing interests The authors declare no competing interests.

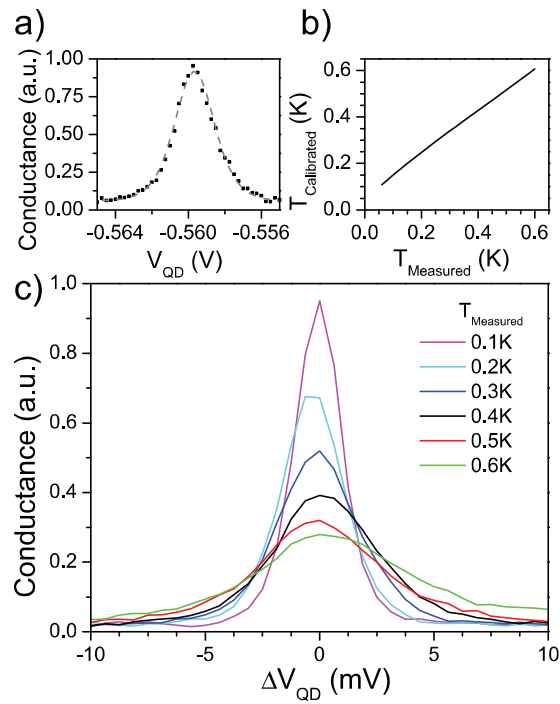
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2058-6>.

Correspondence and requests for materials should be addressed to I.V.B., H.-S.S. or M.Y.

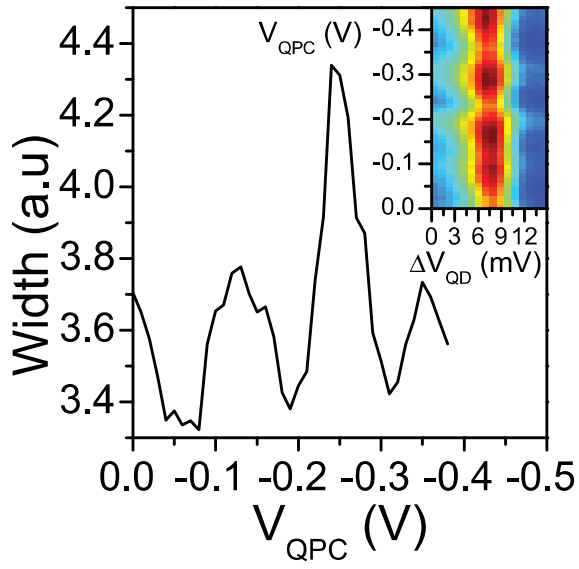
Peer review information *Nature* thanks GuoPing Guo, Robert Peters and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

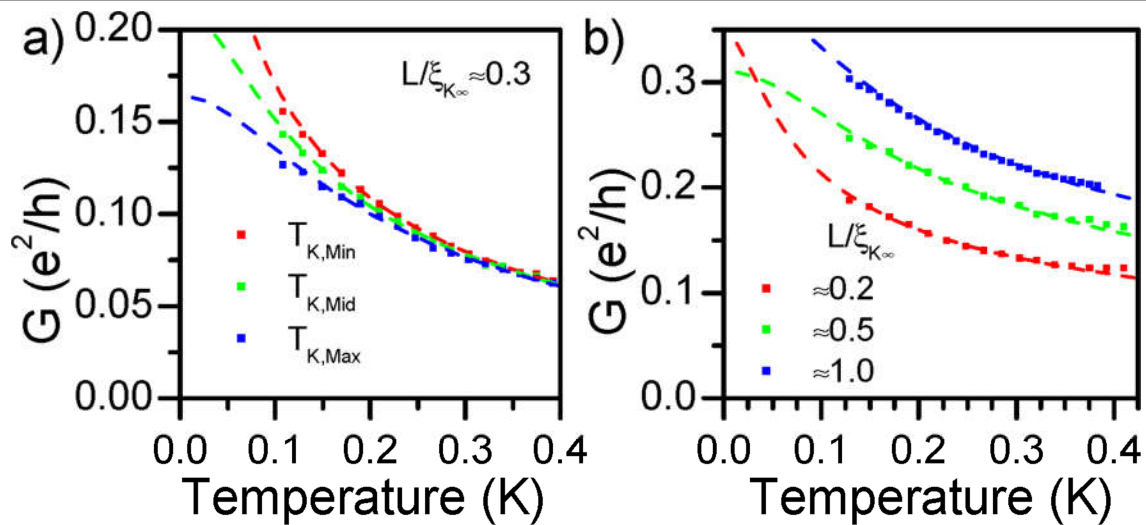


Extended Data Fig. 1 | Calibration of measured sample temperature.

a. Conductance versus QD gate voltage V_G taken at a temperature of 300 mK. The grey lines show the fit of the data to the theoretical lineshape for a Coulomb blockade peak. **b.** The electron temperature in our device $T_{calibrated}$ versus the lattice temperature measured by a thermometer located at the mixing chamber $T_{measured}$. **c.** Conductance of the Coulomb blockade peaks versus gate voltage (around the peak centre) δV_G shown for several values of the measured temperature. (a.u., arbitrary units).



Extended Data Fig. 2 | Full width at half maximum of a Coulomb blockade peak versus QPC gate voltage V_{QPC} . Data are shown for a Coulomb blockade peak away from the Kondo valley. The inset shows conductance G through the QD as a function of (shifted) gate voltage δV_G and QPC gate voltage V_{QPC} . Oscillations of both peak conductance and peak width with respect to V_{QPC} are clearly observed.



Extended Data Fig. 3 | Plot of conductance versus temperature at the centre of the Kondo valley. The lines are fits to the empirical formula. **a**, The data in red correspond to V_{QPC} , at which the minimum Kondo temperature was observed. The data in blue correspond to V_{QPC} , at which the maximum Kondo

temperature was observed. Green data points correspond to data taken at a midpoint between the red and blue datasets. **b**, Data at $V_{QPC} = 0$ V, for several QD settings corresponding to different $L/\xi_{K\infty}$.

Extended Data Table 1 | Coulomb-blockade peak FP oscillation amplitude

QPC distance L	Oscillation amplitude ΔG
1.4 μm	0.080 e^2/h
3.6 μm	0.073 e^2/h
6.1 μm	0.071 e^2/h

The amplitude of FP oscillations in Coulomb-blockade peak conductance versus distance to the QPC gates.

Extended Data Table 2 | Fitting parameters with upper and lower bounds

Fitting Parameter	Lower Bound	Upper Bound
Kondo Temperature T_K	0K	1.2K
Conductance at $T = 0$ G_0	0	e^2/h
Power Law Factor s	0.21	0.23

Current-driven magnetic domain-wall logic

<https://doi.org/10.1038/s41586-020-2061-y>

Received: 24 May 2019

Accepted: 16 January 2020

Published online: 11 March 2020

 Check for updates

Zhaochu Luo^{1,2}, Aleš Hrabec^{1,2,3}, Trong Phuong Dao^{1,2,3}, Giacomo Sala³, Simone Finizio², Junxiao Feng³, Sina Mayr^{1,2}, Jörg Raabe², Pietro Gambardella^{3,4} & Laura J. Heyderman^{1,2,4}

Spin-based logic architectures provide nonvolatile data retention, near-zero leakage, and scalability, extending the technology roadmap beyond complementary metal–oxide–semiconductor logic^{1–13}. Architectures based on magnetic domain walls take advantage of the fast motion, high density, non-volatility and flexible design of domain walls to process and store information^{1,3,14–16}. Such schemes, however, rely on domain-wall manipulation and clocking using an external magnetic field, which limits their implementation in dense, large-scale chips. Here we demonstrate a method for performing all-electric logic operations and cascading using domain-wall racetracks. We exploit the chiral coupling between neighbouring magnetic domains induced by the interfacial Dzyaloshinskii–Moriya interaction^{17–20}, which promotes non-collinear spin alignment, to realize a domain-wall inverter, the essential basic building block in all implementations of Boolean logic. We then fabricate reconfigurable NAND and NOR logic gates, and perform operations with current-induced domain-wall motion. Finally, we cascade several NAND gates to build XOR and full adder gates, demonstrating electrical control of magnetic data and device interconnection in logic circuits. Our work provides a viable platform for scalable all-electric magnetic logic, paving the way for memory-in-logic applications.

Our concept for chiral magnetic domain-wall (DW) logic takes advantage of the efficiency and speed of magnetic DW motion induced by spin–orbit torques (SOTs)^{16,21–27} and exploits the chiral coupling between adjacent magnets with competing magnetic anisotropy and interfacial Dzyaloshinskii–Moriya interaction (DMI)^{17–20} (Fig. 1a). Based on this coupling, we demonstrate that it is possible to invert a DW using an electric current, namely, to transform an up/down (\odot/\otimes) DW into a down/up (\otimes/\odot) DW or vice versa. Here, the magnetization directions of \otimes and \odot in the racetrack represent the Boolean logical values ‘1’ and ‘0’, respectively. In Fig. 1b, we illustrate the design of a DW inverter, which consists of an in-plane (IP) magnetized region embedded in a racetrack with out-of-plane (OOP) magnetization. As the two OOP-magnetization regions (hereafter, ‘OOP regions’) on either side of the IP-magnetization region (hereafter, ‘IP region’) are coupled by the DMI, the reversal of one OOP region induces the reversal of the other, leading to the inversion of a DW travelling along the racetrack, which is equivalent to a NOT gate.

To demonstrate the operation of the DW inverter, we fabricated a set of OOP magnetic Pt/Co/AlO_x nanowires with 50-nm-wide V-shaped IP regions patterned using a selective oxidation process (Fig. 2a, Extended Data Fig. 1). The DW motion is driven by an electric current (Fig. 2b) and tracked with polar magneto-optic Kerr effect (MOKE) microscopy. Starting from the initial down–right–up magnetization configuration of the OOP–IP–OOP structure ($\otimes \rightarrow \odot$), an \odot/\otimes DW is injected from the left OOP region (Fig. 2c). By applying a sequence of current pulses, the \odot/\otimes DW moves in the direction of the current towards the IP region, as expected for a left-handed chiral Néel DW^{22–24}. When the \odot/\otimes DW encounters the IP region, the IP magnetization switches, going from

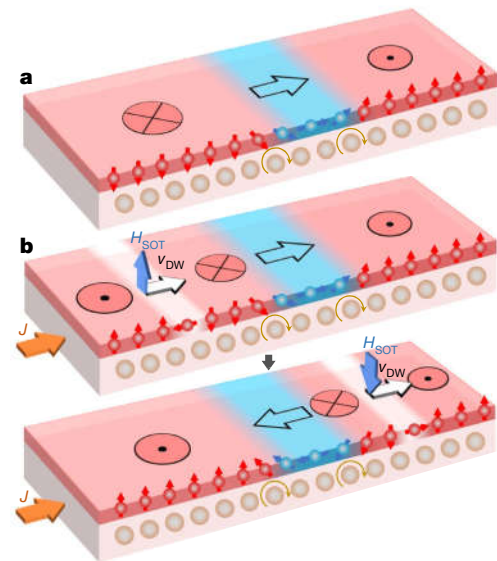


Fig. 1 | Chiral coupling between adjacent nanomagnets and current-driven DW inversion. **a**, Schematic of magnetic chiral coupling induced by the interfacial DMI. After selective oxidation, the magnetizations of neighbouring OOP (oxidized; red-shaded) and IP (unoxidized; blue-shaded) regions align with a left-handed chirality in Pt/Co/AlO_x. **b**, Schematic showing current-driven DW inversion, which occurs as the DW transfers across the IP region. The white-shaded region is the DW, and the directions of the effective field induced by the SOTs, H_{SOT} , and the DW velocity, v_{DW} , are indicated with arrows. In Pt/Co/AlO_x, both \odot/\otimes and \otimes/\odot DWs move in the same direction as the electric current J .

¹Laboratory for Mesoscopic Systems, Department of Materials, ETH Zurich, Zurich, Switzerland. ²Paul Scherrer Institut, Villigen, Switzerland. ³Laboratory for Magnetism and Interface Physics, Department of Materials, ETH Zurich, Zurich, Switzerland. ⁴These authors jointly supervised this work: Pietro Gambardella, Laura J. Heyderman. ✉e-mail: zhaochu.luo@psi.ch; pieter.gambardella@mat.ethz.ch; laura.heyderman@psi.ch

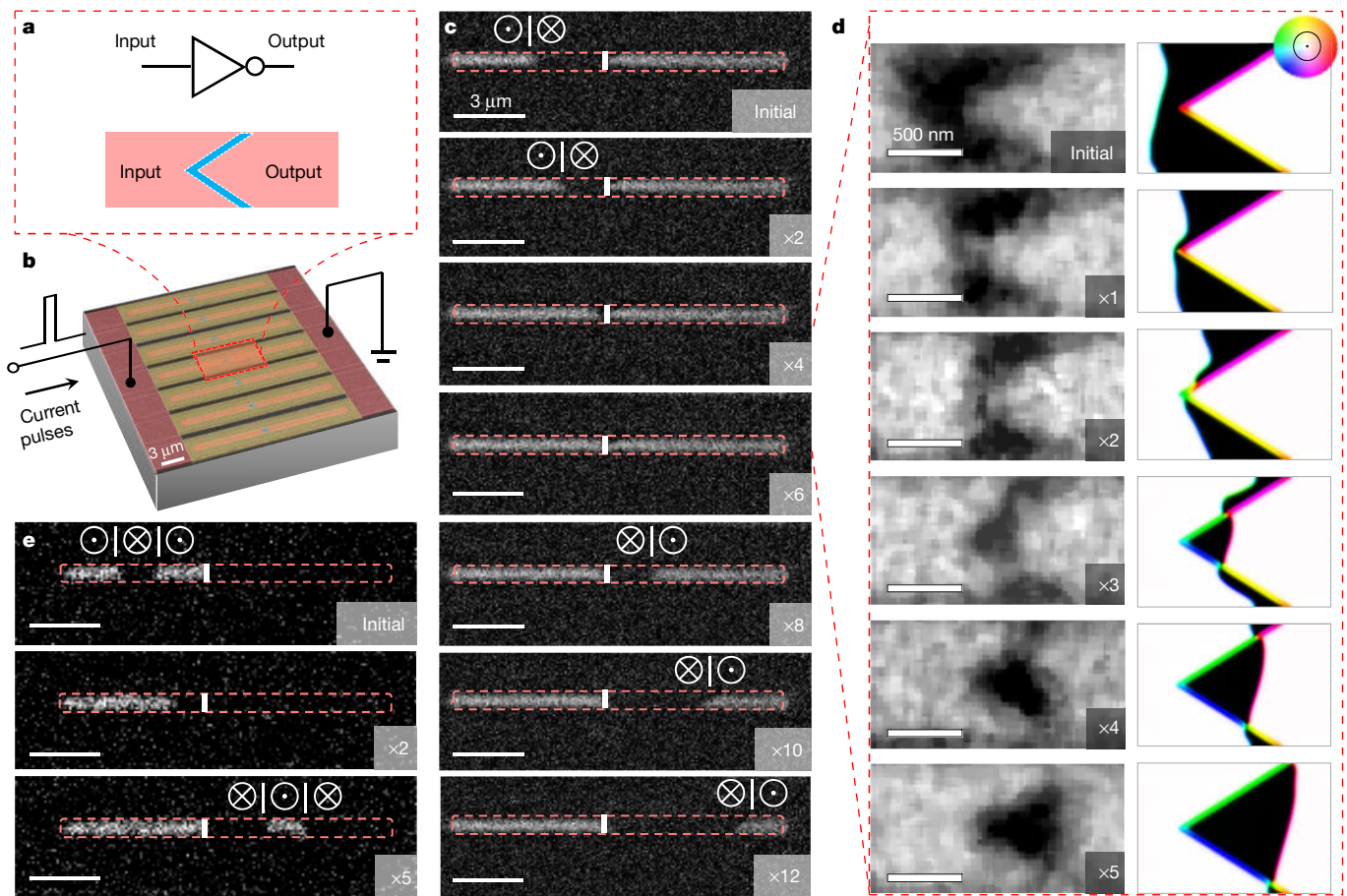


Fig. 2 | Current-driven DW inverter. **a**, Schematics of a NOT gate and current-driven DW inverter. Red- and blue-shaded regions indicate regions with OOP and IP anisotropies, respectively. **b**, Coloured scanning electron microscope (SEM) image of seven parallel DW inverters in a three-dimensional rendering of the DW measurement setup. The direction of the current pulses is indicated. **c**, MOKE image sequence of DW inversion for a DW incident from the left with an $\odot|\otimes$ configuration. The edges of the magnetic racetracks are indicated by red dashed lines and the positions of the inverters are shown by white lines. The bright and dark regions in the racetracks in the MOKE images correspond to \odot and \otimes magnetization, respectively. The entire image sequence for seven inverters is shown in Supplementary Video 1. **d**, X-ray magnetic circular dichroism (XMCD) image sequence of DW inversion for an incident DW with an $\odot|\otimes$ configuration measured by STXM. Each image is captured after the

application of one current pulse. The bright and dark regions in the XMCD images correspond to \odot and \otimes magnetization, respectively. Micromagnetic simulations of the inversion process are shown on the right of each image, with the direction of the magnetization indicated by the colour wheel. The entire image sequence is shown in Supplementary Video 2. **e**, MOKE images showing the inversion of a \otimes domain driven across the IP region with current pulses. The entire image sequence is shown in Supplementary Video 3. The current density and duration of the pulses in **c** and **e** are $7.5 \times 10^{11} \text{ A m}^{-2}$ and 50 ns, whereas in **d** they are $1.1 \times 10^{12} \text{ A m}^{-2}$ and 1 ns. The scale bars are $3 \mu\text{m}$ in the MOKE images and 500 nm in the XMCD images with simulations. In **c–e**, the number of current pulses applied before the acquisition of each image is indicated at the bottom right.

$\otimes \rightarrow \odot$ to $\otimes \leftarrow \odot$, accompanied by the annihilation of the DW to the left of the IP region and nucleation of a new DW with opposite polarity to the right of the IP region. Insight into the microscopic mechanism of the DW inversion is provided by the combination of scanning transmission X-ray microscopy (STXM) and micromagnetic simulations (Fig. 2d, Supplementary Figs. 1, 2). As the incident $\odot|\otimes$ DW approaches the IP region, it is compressed against the IP region by the SOTs, increasing both the magnetostatic and exchange energies. The resulting compact, high-energy spin texture²⁸ can only be unwound by annihilating the incident DW and switching the IP magnetization with the help of SOTs. Upon switching of the IP magnetization from \rightarrow to \leftarrow , a \otimes domain nucleates on the right side of the IP region because of chiral coupling. This process is promoted at the tip of the V-shaped inverter owing to the additive contribution of chiral coupling from both sides of the V-shaped region. Therefore, the optimized design of the narrow V-shaped IP region facilitates the switching of the IP magnetization and the nucleation of a new domain (Extended Data Fig. 2). As a result, the $\odot|\otimes$ DW is effectively transmitted through the IP region and transformed into a

$\otimes|\odot$ DW. An analogous inversion process occurs for an incident $\otimes|\odot$ DW (see Extended Data Fig. 3), so that the inverter effectively reverses the magnetization of domains travelling across the IP region, as shown in Fig. 2e. By using an electric current, it is possible to invert not only a single DW, but also a sequence of DWs, and consequently a sequence of domains that propagate along a racetrack. This is a unique feature of chirally coupled nanomagnetic structures.

Building on the principles used to construct the NOT gate, we now demonstrate how to realize a reconfigurable NAND/NOR gate. This gate makes our concept for current-driven DW logic functionally complete, given that any Boolean function can be implemented using a combination of NAND or NOR gates. The core structure of this gate (indicated by the dashed red circle in Fig. 3a) is composed of four OOP regions (Fig. 3b, in red) that form two logic inputs, one bias and one logic output connected via IP regions (in blue). To illustrate the functionality of the NAND gate, we fabricated four devices with the same core structure and different logic-input configurations (Fig. 3d, e). For each device, two DW reservoirs are connected to the inputs via

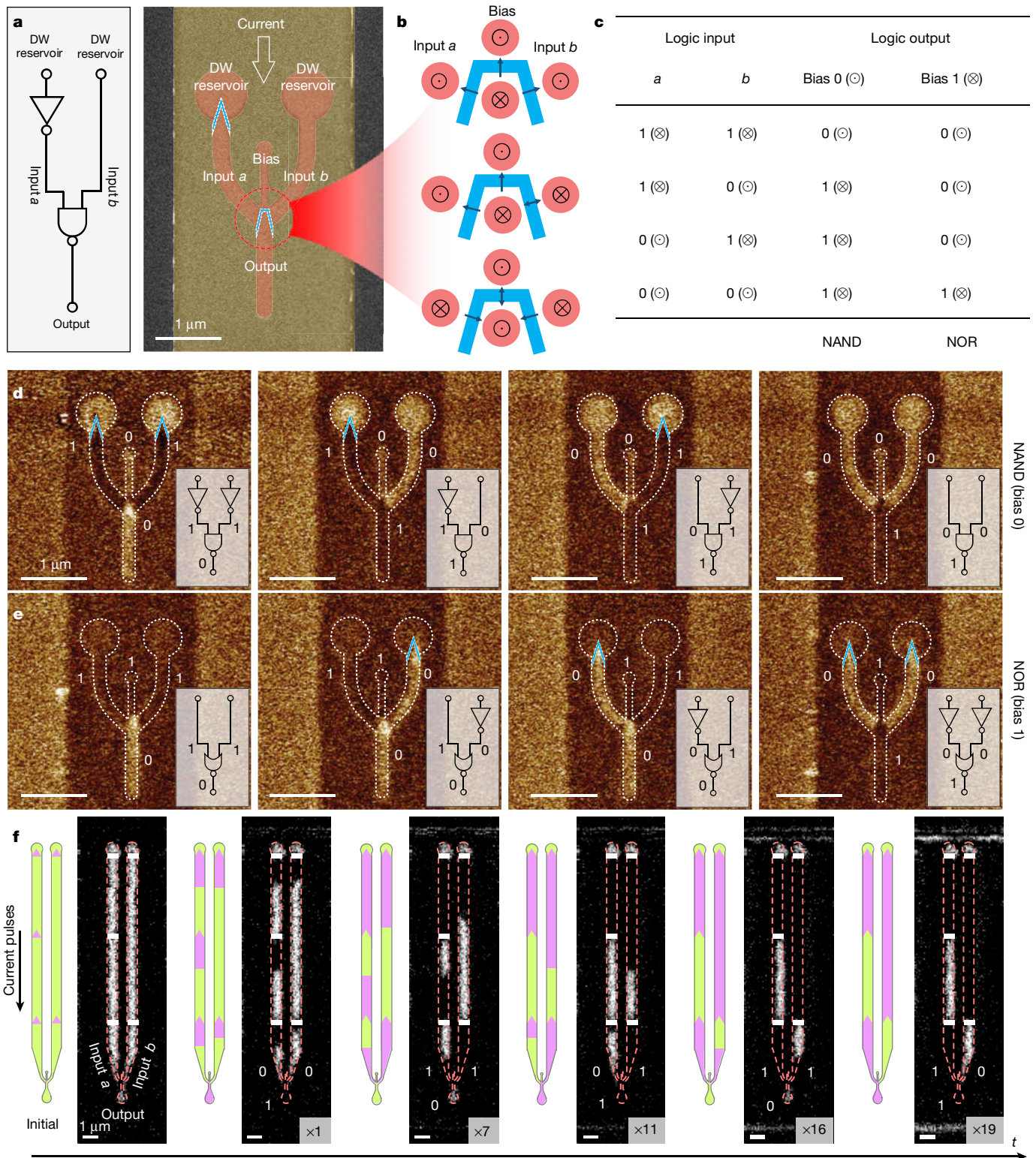


Fig. 3 | Reconfigurable NAND/NOR logic gates. **a**, Coloured SEM image of a reconfigurable NAND/NOR logic gate (right) and corresponding logic-circuit (left). Red, blue and yellow colours indicate regions with OOP and IP magnetic anisotropies and the Pt strip, respectively. **b**, Schematics showing the relationship between the magnetization in the two logic inputs, the bias and the logic output for the NAND gate. **c**, Truth table for the reconfigurable NAND and NOR logic operations. **d**, **e**, MFM images of different logic operations for inputs of '11', '10', '01' and '00' with a bias of '0' (**d**) and '1' (**e**) and the corresponding logic-circuit diagrams. The DW reservoirs (circular pads at the top of each image) and the bias are set to logical value '0' (**d**) and '1' (**e**) by applying an OOP magnetic field of 1 kOe. **f**, MOKE image sequence and

corresponding schematics showing the operation of a single NAND gate with a sequence of logic inputs driven by an electric current. The initial magnetization is set to be ⊙ by applying an OOP magnetic field of 1 kOe. The green and purple areas in the schematics correspond to ⊙ and ⊗ magnetization, respectively. The boundaries of the logic gate are indicated by red dashed lines and the positions of the inverters are indicated by white solid lines in the MOKE images. The direction and the number of current pulses applied before each MOKE image are indicated (current density, $7.5 \times 10^{11} \text{ A m}^{-2}$; pulse length, 30 ns). The entire image sequence is shown in Supplementary Video 5. The bright and dark areas in the device regions in **d–f** correspond to ⊙ and ⊗ magnetization, respectively. All the scale bars are 1 μm.

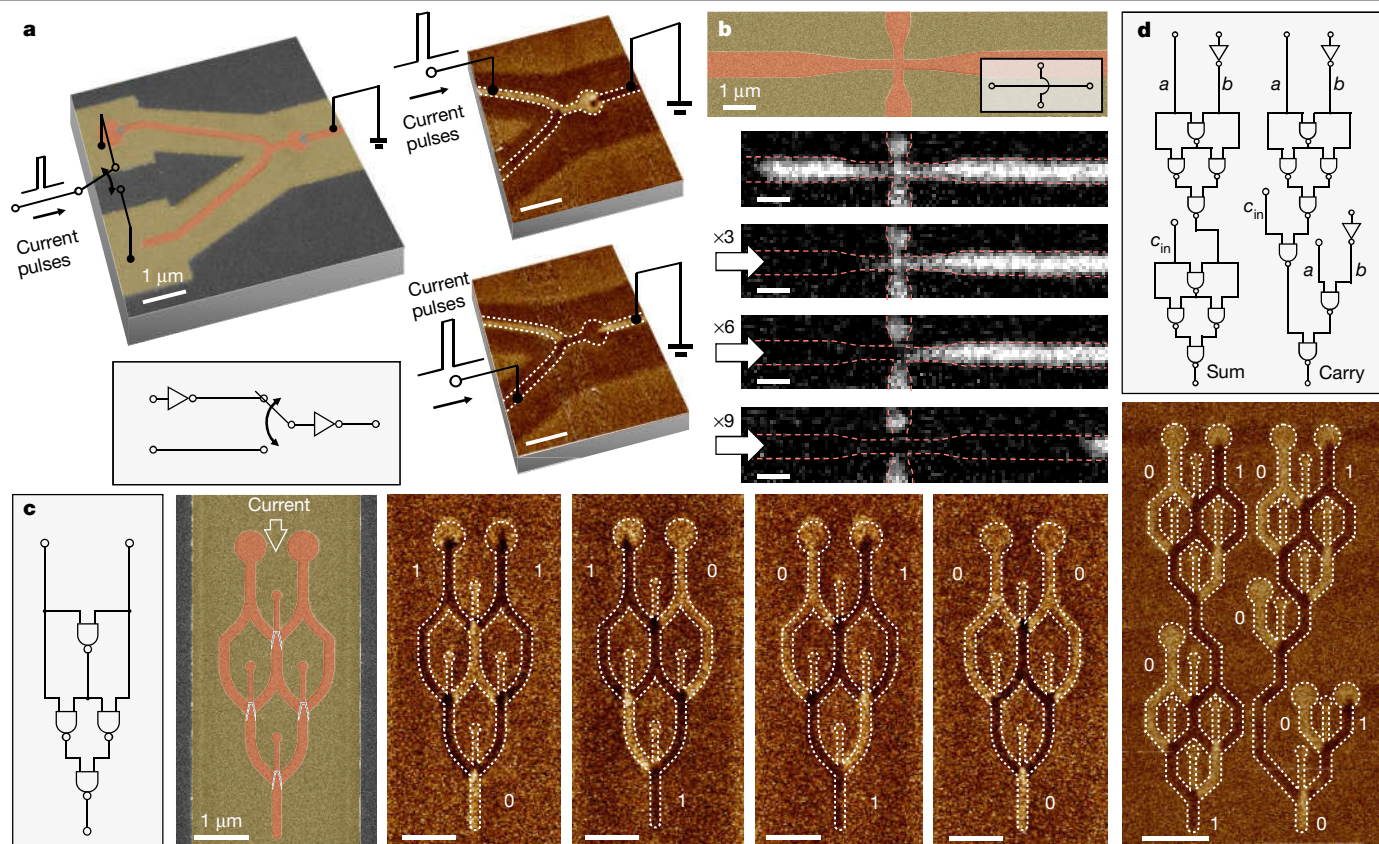


Fig. 4 | Electrical control of data flow and cascaded DW logic circuits. **a**, Data flow switch through a Y-shaped structure, initialized so that the magnetization points \otimes everywhere within the device. Left, coloured SEM image of the device and corresponding logic-circuit diagram. Right, MFM images of the magnetic configuration when current flows through the upper racetrack (top) and then the lower racetrack (bottom). **b**, Sequence of MOKE images illustrating electrical control of data flow through a cross structure. The direction and number of current pulses are indicated (current density $9 \times 10^{11} \text{ A m}^{-2}$, pulse length 30 ns). A coloured SEM image of the cross structure and the corresponding logic-circuit diagram are shown at the top. The entire image sequence is shown in Supplementary Video 6. **c**, XOR gate fabricated by

cascading four NAND gates. Left, coloured SEM image of the XOR logic gate and corresponding logic-circuit diagram. Right, MFM images of the XOR logic gate with logic inputs of '11', '10', '01' and '00'. The DW reservoirs and the bias are set to logical value '0' by applying an OOP magnetic field of 1 kOe. Red, blue and yellow colours in the SEM images indicate the regions with OOP and IP magnetization and the Pt strips, respectively. **d**, Full adder gate. Top, logic-circuit diagram of the full adder with input operands of $a = '0'$, $b = '1'$ and a carry bit of $c_{in} = '1'$. Bottom, MFM image of the full adder magnetic circuit. The bright and dark areas in the device regions in **b–d** correspond to \otimes and \odot magnetization, respectively. All the scale bars are 1 μm .

racetracks. The four logic-input configurations of '11', '10', '01' and '00' are achieved by placing inverters after some of the DW reservoirs. The two DW reservoirs and the bias are set to logical value '0' by applying an OOP magnetic field of 1 kOe. By applying current pulses, the \otimes magnetic domains propagate from the DW reservoirs with inverters (logic inputs '1'), whereas the magnetization stays the same (\odot) for the racetracks connected to the DW reservoirs without inverters (logic inputs '0'). As a result of chiral coupling²⁰, the output magnetization depends on the relative alignments of the inputs and the bias (Fig. 3b, c), which is analogous to a majority gate^{4,8,9}. Therefore, for an \odot bias, the output magnetization switches to \otimes only when both of the input magnetizations are \otimes ; otherwise, the output magnetization is \odot . As shown in the magnetic force microscopy (MFM) images in Fig. 3d, the magnetization direction of the logic output is \odot ('0') for logic inputs of '11' and \otimes ('1') for logic inputs of '10', '01' and '00'. This relationship between the logic inputs and the output corresponds to the required logic operations for a NAND gate (Fig. 3c). By changing the orientation of the bias to \otimes , as shown in Fig. 3e, we can reconfigure the NAND gate into a NOR gate. In the latter case, the output magnetization is \otimes only when both of the input magnetizations are \odot . Hence, our logic gate can be reconfigured between NAND and NOR by switching the bias terminal, enabling rapid logic reprogramming during run time^{2,10}.

We now demonstrate the operation of a single NAND gate using current-driven DW motion to provide a series of different logic inputs to the same gate over time (Fig. 3f). In this device, we place three DW inverters in the left racetrack and two DW inverters in the right racetrack. This means that a sequence of current pulses will generate the sequence of logic inputs: '00'–'11'–'01'–'11'–'10'. The corresponding logic outputs are then '1'–'0'–'1'–'0'–'1' over time. For each operation, the DWs that give the two inputs may not arrive at the gate at the same time. We can mitigate this by introducing a sufficient propagation delay time (see Methods). In a real device, this can be achieved by clocking the electric current. Furthermore, it would be possible to control the logic inputs and bias terminals with magnetic tunnel junctions (MTJs) fabricated on the magnetic racetracks, which can also be used for electrical readout of the outputs²⁹ (see Methods).

In addition to forming a complete logic set, chiral DW racetracks fulfil three additional requirements for practical implementation in logic circuits, namely, input selectivity, data crossover and cascading of different logic gates. Using current-driven DW propagation through the Y-shaped structure in Fig. 4a, it is possible to electrically select the logic input. A simple cross structure allows DWs to propagate in orthogonal racetracks (Fig. 4b, Extended Data Fig. 4), which simplifies the design of crossovers, avoiding the complexity of metal bridges used

in conventional charge-based circuits. Moreover, because our logic inputs and outputs are based on the same physical phenomena, several logic gates can be cascaded directly, without the need for additional transducers between magnetic and electric signals (Extended Data Fig. 5). As examples, we present in Fig. 4c a binary half adder created by cascading four NAND gates to form an XOR gate, and show the full adder operation by cascading 15 NAND gates in Fig. 4d. This circuit also demonstrates the possibility to fan out a single output that can be used to drive the input of the next logic gates. A remaining challenge is to create magnetic logic circuits with feedback loops. This could be realized by using either an external electrical circuit to read the output and write this back to the input with MTJs, or an additional racetrack with an inverted current direction to drive DWs from the output back to the input.

For device applications, the scalability and efficiency of magnetic DW logic circuits should also be addressed. Because the chiral coupling induced by the DMI is effective at the scale of the magnetic moments, it should be possible to reduce the size of the logic gates down to a few nanometres by using advanced lithography techniques. The speed of the logic operations is related to the DW velocity, which can reach several hundreds of metres per second for chiral DWs driven by SOTs^{21–24}. The operation time can be estimated from the time required for a DW to transfer across the gate, which can be as short as a few tens of picoseconds for an inverter scaled down to $10 \times 10 \text{ nm}^2$ (see Methods). The energy consumption of a single NOT operation in our $0.8 \times 1\text{-}\mu\text{m}^2$ racetracks is about 20 pJ, which would scale down to less than 20 aJ in structures with a footprint of $10 \times 10 \text{ nm}^2$ (see Methods). The non-volatility of the magnetic inputs and outputs provides further energy savings because magnetic DW logic circuits do not consume power when idle and do not need reloading of data after power-off. These features make all-electric magnetic DW logic attractive for use in low-power, ‘instant-on’ microelectronic processors that are ubiquitous in modern-day electronics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2061-y>.

1. Allwood, D. A. et al. Submicrometer ferromagnetic NOT gate and shift register. *Science* **296**, 2003–2006 (2002).
2. Ney, A., Pampuch, C., Koch, R. & Ploog, K. H. Programmable computing with a single magnetoresistive element. *Nature* **425**, 485–487 (2003).

3. Allwood, D. A. et al. Magnetic domain-wall logic. *Science* **309**, 1688–1692 (2005).
4. Imre, A. et al. Majority logic gate for magnetic quantum-dot cellular automata. *Science* **311**, 205–208 (2006).
5. Dery, H., Dalal, P., Cywiński, Ł. & Sham, L. J. Spin-based logic in semiconductors for reconfigurable large-scale circuits. *Nature* **447**, 573–576 (2007).
6. Matsunaga, S. et al. Fabrication of a nonvolatile full adder based on logic-in-memory architecture using magnetic tunnel junctions. *Appl. Phys. Express* **1**, 091301 (2008).
7. Behin-Aein, B., Datta, D., Salahuddin, S. & Datta, S. Proposal for an all-spin logic device with built-in memory. *Nat. Nanotechnol.* **5**, 266–270 (2010).
8. Nikonov, D. E., Bourianoff, G. I. & Ghani, T. Proposal of a spin torque majority gate logic. *IEEE Electron Device Lett.* **32**, 1128–1130 (2011).
9. Breikreutz, S. et al. Majority gate for nanomagnetic logic with perpendicular magnetic anisotropy. *IEEE Trans. Magn.* **48**, 4336–4339 (2012).
10. Joo, S. et al. Magnetic-field-controlled reconfigurable semiconductor logic. *Nature* **494**, 72–76 (2013).
11. Chumak, A. V., Serga, A. A. & Hillebrands, B. Magnon transistor for all-magnon data processing. *Nat. Commun.* **5**, 4700 (2014).
12. Curriuan-Incorvia, J. A. et al. Logic circuit prototypes for three-terminal magnetic tunnel junctions with mobile domain walls. *Nat. Commun.* **7**, 10275 (2016).
13. Manipatruni, S. et al. Scalable energy-efficient magnetoelectric spin-orbit logic. *Nature* **565**, 35–42 (2019).
14. Parkin, S. S., Hayashi, M. & Thomas, L. Magnetic domain-wall racetrack memory. *Science* **320**, 190–194 (2008).
15. Franken, J. H., Swagten, H. J. M. & Koopmans, B. Shift registers based on magnetic domain wall ratchets with perpendicular anisotropy. *Nat. Nanotechnol.* **7**, 499–503 (2012).
16. Parkin, S. & Yang, S.-H. Memory on the racetrack. *Nat. Nanotechnol.* **10**, 195–198 (2015).
17. Kubetzka, A., Bode, M., Pietzsch, O. & Wiesendanger, R. Spin-polarized scanning tunneling microscopy with antiferromagnetic probe tips. *Phys. Rev. Lett.* **88**, 057201 (2002).
18. Heide, M., Bihlmayer, G. & Blügel, S. Dzyaloshinskii–Moriya interaction accounting for the orientation of magnetic domains in ultrathin films: Fe/W(110). *Phys. Rev. B* **78**, 140403 (2008).
19. Franken, J. H., Herps, M., Swagten, H. J. & Koopmans, B. Tunable chiral spin texture in magnetic domain-walls. *Sci. Rep.* **4**, 5248 (2014).
20. Luo, Z. et al. Chirally coupled nanomagnets. *Science* **363**, 1435–1439 (2019).
21. Miron, I. M. et al. Fast current-induced domain-wall motion controlled by the Rashba effect. *Nat. Mater.* **10**, 419–423 (2011).
22. Thiaville, A., Rohart, S., Jué, É., Cros, V. & Fert, A. Dynamics of Dzyaloshinskii domain walls in ultrathin magnetic films. *EPL* **100**, 57002 (2012).
23. Ryu, K. S., Thomas, L., Yang, S. H. & Parkin, S. Chiral spin torque at magnetic domain walls. *Nat. Nanotechnol.* **8**, 527–533 (2013).
24. Emori, S., Bauer, U., Ahn, S. M., Martinez, E. & Beach, G. S. Current-driven dynamics of chiral ferromagnetic domain walls. *Nat. Mater.* **12**, 611–616 (2013).
25. Haazen, P. P. J. et al. Domain wall depinning governed by the spin Hall effect. *Nat. Mater.* **12**, 299–303 (2013).
26. Baumgartner, M. et al. Spatially and time-resolved magnetization dynamics driven by spin-orbit torques. *Nat. Nanotechnol.* **12**, 980–986 (2017).
27. Manchon, A. et al. Current-induced spin-orbit torques in ferromagnetic and antiferromagnetic systems. *Rev. Mod. Phys.* **91**, 035004 (2019).
28. Benitez, M. J. et al. Magnetic microscopy and topological stability of homochiral Néel domain walls in a Pt/Co/AlO_x trilayer. *Nat. Commun.* **6**, 8957 (2015).
29. Manfrini, M. et al. Interconnected magnetic tunnel junctions for spin-logic applications. *AIP Adv.* **8**, 055921 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Device fabrication

The magnetic films were deposited on a 200-nm-thick SiN_x layer on a silicon substrate using d.c. magnetron sputtering at a base pressure of $<2 \times 10^{-8}$ torr and a deposition Ar pressure of 3 mtorr, and patterning was carried out by electron-beam lithography. Continuous films of Pt (5 nm)/Co (1.6 nm)/Al (2 nm) were milled into strips with Ar ions through a negative resist (ma-N2401) mask. In these magnetic strips, the upper Co/Al bilayer was milled through a high-resolution positive resist (poly(methyl methacrylate), PMMA) mask to create the DW racetracks and logic devices. To define the IP region in these magnetic structures, a second PMMA mask was patterned by electron-beam lithography on top of the Al layer. Using a low-power (30 W) oxygen plasma at an oxygen pressure of 10 mtorr, the unprotected Al layer was oxidized to induce perpendicular magnetic anisotropy in the Co layer. Finally, electrodes of Cr (5 nm)/Au (50 nm) were fabricated using electron-beam lithography combining electron-beam evaporation with a lift-off process. The main steps of the device fabrication are shown in Extended Data Fig. 1a.

The different anisotropies, with OOP regions (exposed to oxygen plasma) and IP regions (protected by the PMMA mask), were confirmed by polar MOKE measurements (Extended Data Fig. 1b). The effective OOP magnetic anisotropy field was 3.94 kOe, as obtained from anomalous Hall effect measurements with an applied IP magnetic field (Extended Data Fig. 1c). The interfacial DMI constant D was estimated to be $-0.9 \pm 0.1 \text{ mJ m}^{-2}$ (1s.d.) by measuring the DMI-induced chiral coupling²⁰.

Electrical measurement configuration

The magnetic DW motion and logic operation are driven by current pulses generated with an HP Agilent 8114A high-voltage pulse generator and an AVTECH ultrahigh-speed pulse generator. The pulse generators can provide pulses of variable voltage and pulse width. The current densities are calculated by dividing the nominal voltage by the device resistance and cross-sectional area, and are indicated for each operation. The directions of the current pulses for each device are summarized in Extended Data Fig. 6.

MFM measurements

The MFM measurements were performed using a Bruker Dimension Icon Scanning Station mounted on a vibration- and sound-isolation table using tips coated with CoCr. To minimize the influence of the stray field from the MFM tip during the measurements, a thin PMMA layer (~20 nm thick) was spin-coated on the samples to increase the distance between the tip and the magnetic film. The MFM images of the devices were captured after saturation with an OOP magnetic field to set the initial magnetization direction in all of the reservoirs, followed by current pulses applied in the Pt strips to obtain the final states.

MOKE microscopy measurements

The MOKE images were recorded using a custom-built wide-field MOKE microscope. A background image was captured after the application of a large positive OOP magnetic field of 1 kOe. The background image was subtracted from the subsequent images to achieve differential images with magnetic contrast. To prepare the initial state of the DWs shown in Fig. 2c, the racetrack was first saturated with an OOP magnetic field. The magnetic field was removed, leaving the racetrack with OOP magnetization, with a small area of reversed magnetization enclosed by the V-shaped IP region that results from the chiral coupling²⁰. Then, current pulses were applied in the opposite direction to that shown in Fig. 2b to create a single DW on the left side of the DW inverter (Fig. 2c).

STXM measurements

The magnetic configuration of the DW inverter was imaged using STXM at the PoLLux beamline of the Swiss Light Source. The magnetization

state was probed using XMCD at the Co L₃ absorption edge at normal incidence. The devices measured using STXM were fabricated on X-ray-transparent SiN_x membranes.

Micromagnetic simulations

To understand the mechanism of the DW inversion, micromagnetic simulations were carried out with the MuMax3 code³⁰ using a computation box containing $2,048 \times 1,024 \times 1$ cells with $2 \times 2 \times 1.6 \text{ nm}^3$ discretization and the following magnetic parameters: saturation magnetization $M_s = 0.9 \text{ MA m}^{-1}$, effective OOP anisotropy field $H_{\text{eff}} = 150 \text{ mT}$, exchange constant $A = 16 \text{ pJ m}^{-1}$, spin Hall angle of Pt $\theta_{\text{sh}} = 0.1$ and interfacial DMI constant $D = -1.5 \text{ mJ m}^{-2}$.

Mechanism for DW inversion

To elucidate the basic mechanism behind DW inversion in an OOP–IP–OOP structure, we consider a simple model. The DW inversion process can be explained in terms of the effective DMI field generated in non-collinear magnets, where the DMI vector lies in the plane of the magnetic thin film. This effective DMI field is given by:

$$\mathbf{H}_{\text{DMI}} = \frac{2D}{\mu_0 M_s} \left(\frac{\partial m_z}{\partial x}, 0, -\frac{\partial m_x}{\partial x} \right) \quad (1)$$

where μ_0 is the vacuum permeability, and m_z and m_x are the z and x components of the magnetization, respectively. We can then consider a situation in which an $\odot \otimes$ DW is driven by SOTs towards the IP-magnetized region, as shown in Supplementary Fig. 1a. At equilibrium, the IP-magnetized region, together with the surrounding domains, forms a $\otimes \rightarrow \odot$ configuration, which is stabilized by the DMI fields (pointing along $+x$) indicated by $\mathbf{H}_{\text{DMI}}(\text{IP})$. On applying an electric current, the magnetization is subject to an effective field \mathbf{H}_{SOT} induced by the SOTs, which is given by:

$$\mathbf{H}_{\text{SOT}} = \frac{\hbar \theta_{\text{SH}} J}{2\mu_0 e M_s t} \mathbf{m} \times \boldsymbol{\sigma} \quad (2)$$

where \hbar , θ_{SH} , J , e , M_s , t , \mathbf{m} and $\boldsymbol{\sigma}$ are the reduced Planck constant, spin Hall angle, electric current density, electron charge, saturation magnetization, thickness of the magnetic layer, direction of magnetization and direction of the spin polarization at the Pt/Co interface, respectively. Owing to the chiral coupling, the magnetization in the middle of the $\odot \otimes$ DW points along $-x$ (in blue). As shown in Supplementary Fig. 1b, $\mathbf{H}_{\text{SOT}}(\text{DW})$ points to $+z$, so the $\odot \otimes$ DW will propagate along the current direction. As soon as the $\odot \otimes$ DW approaches the IP region, the magnetization in the IP region experiences a dipolar field $\mathbf{H}_{\text{dip}}(\text{IP})$ generated by the IP magnetization of the $\odot \otimes$ DW that points along $-x$. The SOTs also compress the incident DW against the IP region, increasing the DW energy. This results in a compact, high-energy spin texture containing two closely spaced regions with tail-to-tail IP magnetization, as shown with the associated magnetic charges in Supplementary Fig. 1b: one IP region is in the middle of the $\odot \otimes$ DW with magnetization pointing along $-x$ (in blue), and the other IP region is in the inverter and has magnetization pointing along $+x$ (in red). At a certain point in time, the dipolar field becomes strong enough to switch the magnetization in the IP region from $+x$ to $-x$ with the help of SOTs (Supplementary Fig. 1b, c). Simultaneously, the high-energy spin texture on the left side of the IP region collapses, bringing about the annihilation of the \otimes domain on the left side of the IP region, shown in grey in Supplementary Fig. 1b. After the reversal of the magnetization in the IP region, a reversed \otimes domain is nucleated on the right side of the IP region (shown in grey in Supplementary Fig. 1c), which is a consequence of $\mathbf{H}_{\text{DMI}}(\text{OOP})$ pointing along $-z$. The magnetization points along $+x$ in the middle of the resulting $\otimes \odot$ DW and $\mathbf{H}_{\text{SOT}}(\text{DW})$ points along $-z$ so that this new DW is then transported by the electric current towards $+x$ (Supplementary Fig. 1c).

Whereas this simple model provides insight into the mechanism of DW inversion, the detailed magnetization dynamics is more complex. We have therefore performed micromagnetic simulations. Here, an $\odot\otimes$ DW is driven by an electric current with current density $3 \times 10^{12} \text{ A m}^{-2}$ in a narrow wire containing a straight, 30-nm-wide IP region (Supplementary Fig. 2a). The OOP anisotropy field is set to $H_k = 1.5 \text{ kOe}$ in the OOP region and to zero in the IP region. All three components of the magnetization are recorded at three different positions along the wire: in the centre of the IP region and 30 nm away from the centre on each side (see the dots in Supplementary Fig. 2a). We show in Supplementary Fig. 2b–d how the magnetization responds to the approaching DW. As the $\odot\otimes$ DW approaches the left side of the IP region, the magnetization on the left side of the IP region reverses from $-z$ to $+z$ (see Supplementary Fig. 2b). The magnetization in the IP region reverses from $+x$ to $-x$ along the path shown in Supplementary Fig. 2c to reduce the energy associated with the accumulated magnetostatic charges (shown schematically in Supplementary Fig. 1b). The magnetization on the right side of the IP region is then forced to switch from $+z$ to $-z$ by the chiral coupling (see Supplementary Fig. 2d).

The DMI is critical in the realization of current-driven DW inversion, not only to achieve current-driven DW motion, but also because of its role in the nucleation of the reversed domain. The role of DMI in current-driven DW motion has been studied elsewhere^{22–24}. Here, we determine the role of DMI in the DW inversion process using micromagnetic simulations by varying the DMI value and the OOP anisotropy in the IP region. The DMI–OOP anisotropy phase diagram for current-driven DW inversion is shown in Supplementary Fig. 3 for a current density of $3 \times 10^{12} \text{ A m}^{-2}$ in a narrow wire containing a straight, 30-nm-wide IP region. For zero OOP anisotropy in the IP region, the DW can be inverted when $|D| > 1 \text{ mJ m}^{-2}$. If the DMI is reduced, it does not provide sufficient chiral coupling to nucleate the reverse domain, so the incident DW cannot be inverted. By introducing OOP anisotropy into the IP region, which is expected from a Pt/Co interface, the energy for the DW inversion is reduced and the DMI operational window increases.

To verify the impact of the IP width on the DW inversion process, we performed additional micromagnetic simulations of the magnetization dynamics in the inverter for various widths of the IP region. The outcomes of the simulations for a current density of $3 \times 10^{12} \text{ A m}^{-2}$ and $D = -1.5 \text{ mJ m}^{-2}$ are given in Supplementary Table 1. In the table, a tick indicates that an inverted DW propagates from the IP region into the OOP region as required. If the IP region is too narrow ($< 25 \text{ nm}$), the OOP regions on either side of the inverter are strongly coupled antiparallel²⁰ and the SOTs induced by the current are not strong enough to overcome the chiral coupling. If the width of the IP region is too large ($> 35 \text{ nm}$), the chiral coupling becomes too weak to induce antiparallel coupling of the OOP magnetizations on the left and right sides of the IP magnetization²⁰. The DW is then simply annihilated in the IP region, without any further magnetization dynamics occurring on the other side of the inverter. The results of the micromagnetic simulations were confirmed by experiment: for a straight DW inverter in an 800-nm-wide racetrack, the DW was successfully transferred across a 50-nm-wide DW inverter, but not across a 100-nm-wide inverter. In addition, as shown in Supplementary Table 1, it is possible to increase the operational window of the IP region by including a small OOP anisotropy in the IP region.

Influence of the shape of the IP region of the DW inverter

Here we describe experiments comparing the performance of straight and V-shaped DW inverters with a 50-nm-wide IP region, beginning with the measurements of the straight IP inverter. As shown in the STXM images in Extended Data Fig. 2a, when the $\otimes\odot$ DW encounters the IP region, it annihilates to the left of the IP region, and a new DW with opposite polarity will be nucleated to the right of the IP region. We performed several inversion operations in the same inverter with a straight IP region and found that the reversed magnetic domains nucleate at different locations (Extended Data Fig. 2b). This implies

that the nucleation of the reversed magnetic domain is assisted by random thermal fluctuations or local inhomogeneities.

To improve the reliability of the DW inverter, we implemented a V-shaped IP region, which has two main advantages: first, the tip of the V shape offers a convenient nucleation site for the reversed magnetic domain. This is because, at the tip of the V shape, the output OOP region is surrounded by the input OOP region and experiences the strongest antiparallel chiral coupling. In the STXM measurements, we found that the nucleation of the reversed magnetic domain was located at the tip of the V shape for five out of five operations. Second, the V shape of the IP region leads to lower magnetostatic energy, thus lowering the energy barrier for DW inversion. As shown in Extended Data Fig. 2c, we compared the effective DW velocity measured in two DW inverters, one with a V-shaped and one with a straight IP region (see the method outlined in the Methods section ‘Estimation of the speed of a logic operation’). The velocity of the DW transferring across the V-shaped IP region is higher than that in the straight IP region, and the standard deviation of the velocity is smaller in the inverter with the V-shaped IP region than in the inverter with the straight IP region. This demonstrates the higher efficiency and reliability of the V-shaped IP region as a DW inverter.

Estimation of the speed of a logic operation

Here we describe the estimation of the speed of a logic operation in the NOT gates. First, we measure the DW velocity, v_{DW} , in the uniform OOP region of the racetracks. Then we determine the DW displacement, L_{DW} , from S_1 to S_2 across the NOT gate, following N current pulses (see schematic in Extended Data Fig. 7). From this, we can obtain the time taken by the DW to transfer across the NOT gate, t_{NOT} , and therefore the effective velocity of the DW, v_{NOT} , as it transfers across the NOT gate and is inverted:

$$t_{\text{NOT}} = N t_{\text{pulse}} - \frac{L_{\text{DW}} - L_{\text{NOT}}}{v_{\text{DW}}} \quad (3)$$

$$v_{\text{NOT}} = \frac{L_{\text{NOT}}}{t_{\text{NOT}}}$$

where t_{pulse} is the duration of one current pulse and L_{NOT} is the length of the NOT gate. With this method, we can determine v_{DW} and v_{NOT} as a function of current density (Extended Data Fig. 7). The pulse length was decreased to 2 ns for high current densities to reduce heating (inset in Extended Data Fig. 7). We find that the velocity of the DW in the NOT gate can reach $160 \pm 17 \text{ m s}^{-1}$ for a current density of $1.65 \times 10^{12} \text{ A m}^{-2}$. This value of the DW velocity is used to estimate the energy consumption (see main text).

Here we estimate the time required for the DW to transfer across the DW inverter with the dimensions indicated in Extended Data Fig. 6, scaled down to $10 \times 10 \text{ nm}^2$. Taking the effective inverter DW velocity determined from the experiment, 160 m s^{-1} , the time for a DW to transfer across the downscaled inverter is $\sim 60 \text{ ps}$. For a more accurate estimation, we perform micromagnetic simulations for a downscaled inverter with dimensions of $10 \times 10 \text{ nm}^2$. At such a small scale, the device design is limited by the feature size that can be nanofabricated. Therefore, a straight IP region is considered instead of a V-shaped one, with a width of 10 nm. Taking the simulated effective inverter DW velocity of 118 m s^{-1} , the time for a DW to transfer across the inverter is 85 ps, which is similar to the rough estimation above. The speed of operation can be further improved by optimization of the material to increase the DW velocity—for example, by using an amorphous magnetic material such as CoFeB instead of Co—and optimization of the device design.

Using a similar method to that used for the DW inverter, we determined the speed of a logic operation in the NAND gate experimentally. For this, the operation of a NAND gate with two DW inverters in the input reservoirs was captured using MOKE imaging. Following the application of current pulses, DWs propagate through the NAND gate

and perform logic operations (see MOKE images and corresponding schematics in Extended Data Fig. 8). From these images, we determine the time required for the DW to transfer across the NAND gate to be $t_{\text{NAND}} = 74.1 \text{ ns}$ with an effective DW velocity of $v_{\text{NAND}} = 10.8 \text{ m s}^{-1}$ at a current density of $7.5 \times 10^{11} \text{ A m}^{-2}$.

Synchronization and propagation delay times in the DW circuits

Owing to the presence of defects that lead to pinning of the DWs and the intrinsic stochastic nature of current-driven DW motion, the arrival time of DWs at the logic gates can be different. In electronic logic circuits, this is commonly addressed by introducing a propagation delay time for each operation, that is, the circuits are cycled at a rate that is slower than the longest internal propagation delay times. The same concept of propagation delay time can be applied to our magnetic DW logic gates to obtain a stable output that is independent of the arrival time of the input domains. In other words, with sufficient propagation delay time, all the DWs will arrive at the logic gate, which will result in the correct output for a given logic operation.

To demonstrate how the introduction of a propagation delay time can improve the operational reliability of the gate, we consider the simplest case of a NAND gate with the logic inputs changing from '00' to '11' over time. As schematically shown in Extended Data Fig. 9a, the arrival times of $\otimes|\odot$ DWs for logic inputs a and b are different. In this case, the change from '0' to '1' of input a is at a later time than that for input b , that is, $t_a > t_b$, where t_i ($i = a, b$) is the time at which the inputs change. Once both DWs arrive at the gate, it will take some time for the nucleation of the reversed magnetic domain to occur that depends on the effective DW velocity in the NAND gate. The time at which the correct magnetic domain propagates into the output racetrack is defined as the required propagation delay time, t_{delay} . We fabricated NAND gates with '11' logic inputs with different input racetrack lengths to obtain different arrival times for the two logic inputs. On application of current pulses, a $\otimes|\odot$ DW propagates in both the left and the right input racetracks, thus providing a means to test the reliability of the propagation delay time of the logic gate. As shown in Extended Data Fig. 9b, all devices give the correct output '0', demonstrating that the output of the magnetic DW logic gate is independent of the difference in the arrival time of the input DWs.

In general, for all logic operations in the NAND gate, the operation includes: (i) DW propagation in the input racetracks, (ii) transfer of DWs across the logic gate and (iii) DW propagation in the output racetrack. The total operation time, and therefore the required propagation delay time, can then be expressed as:

$$t_{\text{delay}} = \frac{L_{\text{input}} + L_{\text{output}}}{v_{\text{DW}}} + \frac{L_{\text{NAND}}}{v_{\text{NAND}}} \quad (4)$$

where L_{input} , L_{out} and L_{NAND} are the lengths of the input racetrack, output racetrack and NAND gate, respectively. v_{DW} and v_{NAND} are the DW velocities in the magnetic racetrack and in the NAND gate, correspondingly. Assuming that the DW velocities in the magnetic racetrack and NAND gate have a normal distribution:

$$\begin{aligned} v_{\text{DW}} &\approx \mathbf{N}(\bar{v}_{\text{DW}}, \sigma_{\text{DW}}^2) \\ v_{\text{NAND}} &\approx \mathbf{N}(\bar{v}_{\text{NAND}}, \sigma_{\text{NAND}}^2) \end{aligned} \quad (5)$$

with the average velocity of the DW motion in the racetracks (NAND gate) \bar{v}_{DW} (\bar{v}_{NAND}) and the standard deviation σ_{DW} (σ_{NAND}). To obtain a 97.5% probability that the logic operation is successful (see Extended Data Fig. 9c), the required propagation delay time is estimated to be:

$$t_{\text{delay}, 97.5\%} = \frac{L_{\text{input}} + L_{\text{output}}}{\bar{v}_{\text{DW}} - 2\sigma_{\text{DW}}} + \frac{L_{\text{NAND}}}{\bar{v}_{\text{NAND}} - 2\sigma_{\text{NAND}}} \quad (6)$$

To demonstrate that a sufficient propagation delay time can improve the reliability for a statistically significant number of operations in a NAND gate, we placed the output of a NAND gate on a Hall cross (Extended Data Fig. 10a) and performed 1,172 measurements. For each measurement, the NAND gate was saturated with an OOP magnetic field to set the initial magnetization direction to \odot in all of the reservoirs, and a series of current pulses was applied. The output was measured via the anomalous Hall effect on application of a d.c. current. The pulse source and d.c. source were separated by a bias tee. As indicated by the change of the Hall resistance in Extended Data Fig. 10b, the output changed from '0' to '1' and back to '0'. The electrical measurements were verified by MOKE measurements performed on the NAND gate (see MOKE images in Extended Data Fig. 10b). By using 14 current pulses (equivalent to a propagation delay time of $14 \times 30 \text{ ns}$), the reliability of the NAND gate increased to $>95\%$ (Extended Data Fig. 10c).

For a cascaded logic circuit, the propagation delay time is determined by the longest DW propagation route in the circuit. To decrease the propagation delay time, several possible approaches can be employed; for example, scaling down the dimensions of the circuit, increasing the DW velocity and decreasing the DW pinning by optimizing the materials and the nanofabrication to reduce imperfections.

Reliability of the logic gates

To realize large-scale implementation of the logic gates, reliable operation is essential. Here, we evaluate the reliability of the two basic NOT and NAND gates in terms of device-to-device reliability and operational reliability (Supplementary Table 2).

To demonstrate the high device-to-device reliability of the NOT gate, we fabricated 35 NOT gates, and 34 of them (97%) showed successful operation. We also performed the NOT operations with various current densities in the range $4 \times 10^{11} \text{ A m}^{-2}$ to $1.65 \times 10^{12} \text{ A m}^{-2}$ in a single device with 100% operational reliability.

To test the device-to-device reliability of the NAND gate, we fabricated 56 NAND gates with different logic inputs, and the average success rate was found to be 42/56 (75%). The failure of some of the devices may be related to pinning of the DWs by defects in the material or by irregular features resulting from the nanofabrication. In particular, the width of the magnetic racetracks in the NAND gate is 200 nm, compared to 800 nm for the NOT gate, which means that edge roughness can induce additional pinning. We performed 20 operations for each of four selected devices (Supplementary Fig. 4), and they all gave correct outputs, showing high operational reliability.

We now consider the distribution of the device-to-device reliability for different logic inputs. For the 56 NAND gates, we fabricated 14 of each type with logic inputs '00', '11', '01' and '10'. The ratio of the number of NAND gates that give correct outputs to the total number of NAND gates are 13/14, 11/14, 10/14 and 8/14 for logic inputs of '00', '11', '01' and '10', respectively. The device-to-device reliabilities for '00' and '11' inputs are slightly higher than those of the '01' and '10' inputs. This can be understood by considering the energy difference between the '1' and '0' outputs for the '00', '11', '01' and '10' inputs, which is given by:

$$\begin{aligned} \Delta E_{1/0}^{00} &= -(2E_{\text{input}} + E_{\text{bias}}) \\ \Delta E_{1/0}^{11} &= 2E_{\text{input}} - E_{\text{bias}} \\ \Delta E_{1/0}^{01} &= -E_{\text{bias}} \\ \Delta E_{1/0}^{10} &= -E_{\text{bias}} \end{aligned} \quad (7)$$

where $\Delta E_{1/0}^{ij}$ is the energy difference between the '1' and '0' outputs for input ij ($ij = 11, 00, 01$ or 10), E_{input} is the coupling strength between output and input, and E_{bias} is the coupling strength between the output and the bias. From this set of equations, it follows that the stable output for the '11' input is '0' ($\Delta E_{1/0} > 0$) and the stable outputs for inputs '00', '01' and '10' are all '1' ($\Delta E_{1/0} < 0$), which satisfies the truth table of the

NAND operation. For the NAND gate used in the experiment, the bias is slightly smaller than that of the inputs. Because the energy of the coupling between two OOP magnetizations separated by the IP region is proportional to the length of their boundary, the coupling energy between the output and input magnetization is larger than that between the output and bias magnetization, that is, $E_{\text{input}} > E_{\text{bias}}$. Hence, $|\Delta E^{00}| > |\Delta E^{11}| > |\Delta E^{01}| = |\Delta E^{10}|$ for the NAND operation. This trend in the energy difference between the correct and erroneous outputs for different logic inputs correlates well with the trend in the device-to-device reliability for different logic inputs.

We also tested the operational reliability of the cascaded logic circuit (full adder) shown in Fig. 4d. The ratio of the number of successful operations to the total number of operations performed was found to be 28/30.

Therefore, in our proof-of-concept experiments, we have demonstrated the high reliability of the magnetic gates. We further emphasize that there is still room to improve the device-to-device reliability in terms of optimization of the fabrication process, device design and material properties.

Electrical control of logic inputs and detection of logic output

For the proof-of-concept experiment shown in Fig. 3f, we ensured that specific DWs reached the inputs by placing inverters on the input race-tracks. After saturation with an OOP magnetic field, the magnetization direction in the racetrack was set to \odot . On application of a current, the magnetization of the propagating DWs was reversed as they were transferred across each inverter. By placing different numbers of inverters at different positions in the input racetracks, we generated a sequence of logic inputs in order to obtain different inputs at the same gate over time, thus demonstrating the real-time operation of the gate. We also showed that electrical switching of DW propagation in a Y-shaped structure could be used to inject DWs and define specific logic inputs (Fig. 4a).

For downscaled logic circuits, magnetic tunnelling junctions (MTJs) fabricated on the logic-input racetracks would provide a more compact method to control the logic inputs (Supplementary Fig. 5). Indeed, it has been shown that an MTJ on a magnetic racetrack can be used to write magnetic domains via spin transfer torque²⁹. Therefore, MTJs fabricated on magnetic racetracks can be used to electrically control the logic inputs. For the detection of the logic outputs, we used MFM, MOKE microscopy and Hall measurements in our proof-of-concept experiment. For the downscaled logic circuits, MTJs fabricated on the output racetracks could be used to electrically detect the logic outputs.

Moreover, it is practical to not only read the output of a gate but to also transfer it to the input of another gate using MTJ devices, in order to obtain information feedback. The feedback is critical for sequential logic operations such as those performed in a flip-flop gate. The MTJ/racetrack hybrid structure can provide a compact method to perform complex logic.

Energy consumption of downscaled logic devices

To estimate the energy consumption of the inverter used in the experiments, we consider the area containing the V-shaped IP region, where the DW is reversed. The energy consumption per operation of the inverter is calculated from the power-delay product in the bottom Pt layer:

$$E = I^2 R t = \frac{J^2 \rho W L^2 h}{v_{\text{NOT}}} \quad (8)$$

where J , ρ , W , L , h and v_{NOT} represent the current density, resistivity of Pt, inverter width, inverter length, thickness of Pt layer (5 nm) and effective DW velocity in the inverter, respectively. Taking the inverter dimensions

($W \times L = 0.8 \times 1.0 \mu\text{m}^2$), the resistivity of a thin Pt film ($\rho = 30.0 \mu\Omega \text{ cm}$), the experimentally measured current density ($J = 1.65 \times 10^{12} \text{ A m}^{-2}$) and the effective inverter DW velocity ($v_{\text{NOT}} = 160 \text{ m s}^{-1}$), the energy consumption per operation of the inverter is calculated from equation (8) to be 20.4 pJ.

For a rough estimation of the energy consumption of a downscaled inverter, we scale down the dimensions of the inverter while keeping the value of the Pt layer thickness, the Pt resistivity, the current density and the effective DW velocity across the inverter the same as those measured in the experiment. The energy consumption per operation for an inverter with the dimensions indicated in Extended Data Fig. 2 scaled down to $10 \times 10 \text{ nm}^2$ is 25.5 aJ. For a more accurate estimation, we performed micromagnetic simulations for a downscaled inverter with lateral dimensions of $10 \times 10 \text{ nm}^2$. At such a small scale, the device design is limited by the feature size that can be nanofabricated. Therefore, a straight IP region is considered instead of a V-shaped one, with a width of 10 nm. Using a simulated current density of $1.2 \times 10^{12} \text{ A m}^{-2}$ and effective inverter DW velocity of 118 m s^{-1} in equation (8), we find that the energy consumption per operation is 18.4 aJ, which is similar to the rough estimate above. This energy consumption for the downscaled inverter is comparable to the switching energy of $\sim 30 \text{ aJ}$ found in advanced complementary metal-oxide-semiconductor devices¹³. Further improvement of the energy consumption can be achieved by optimizing the material and device design in order to decrease the required current density and increase the DW velocity.

The above estimation concerns only the energy consumed in the logic gate, that is, the energy consumed for the inversion of a DW. Additional energy is required to nucleate the DWs in the racetrack for the logic inputs, to detect the logic outputs and to move DWs along the interconnections. The total energy consumption therefore depends on the detailed design of the logic circuit.

Data availability

All data used in this paper have been deposited in the Zenodo database, at <https://doi.org/10.5281/zenodo.3557288>.

30. Vansteenkiste, A. et al. The design and verification of MuMax3. *AIP Adv.* **4**, 107133 (2014).

Acknowledgements We thank A. Weber, V. Guzenko and X. Wang for technical support with sample fabrication and measurements. This work was supported by the Swiss National Science Foundation through grant number 200020-172775. S.M. acknowledges funding from the Swiss National Science Foundation under grant agreement number 200021-172517. A.H. was funded by the European Union's Horizon 2020 research and innovation programme through Marie Skłodowska-Curie grant agreement number 794207 (ASIQS). J.F. was partially supported by a fellowship from the Chinese Scholarship Council. Part of this work was performed at the PolLux (X07DA) endstation of the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. The PolLux endstation was financed by the German Bundesministerium für Bildung und Forschung under grant agreements 05KS4WE1/6 and 05KS7WE1. Part of this work was performed at the Scanning Probe Microscopy Laboratory, Laboratory for Micro and Nanotechnology, Paul Scherrer Institut, Villigen, Switzerland.

Author contributions Z.L., L.J.H. and P.G. conceived the work and designed the experiments; Z.L. fabricated the samples and performed the MFM and MOKE measurements with the support of A.H., T.P.D. and J.F.; Z.L. analysed and interpreted the data from the MOKE measurements with the help of A.H., T.P.D. and P.G.; Z.L., A.H., G.S., S.F., T.P.D., J.F., S.M. and J.R. performed the STXM measurements and interpreted the data; A.H. performed the micromagnetic simulations; Z.L., P.G. and L.J.H. worked on the manuscript together. All authors contributed to the discussion of the results and the manuscript revision.

Competing interests Z.L., A.H., T.P.D., P.G. and L.J.H. have filed a European patent application (EPO) covering the logic architectures based on current-driven domain-wall motion.

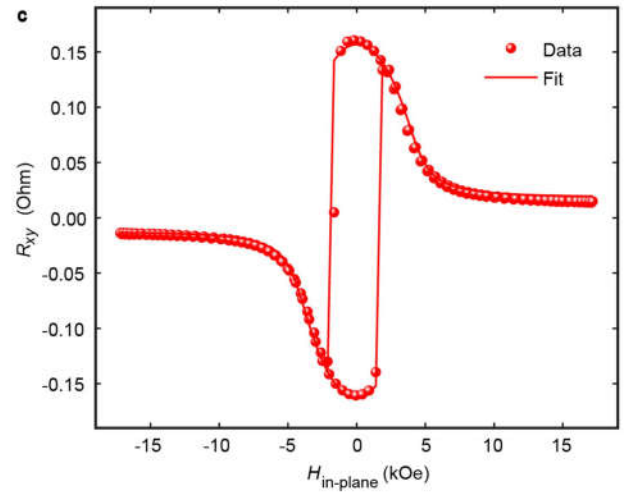
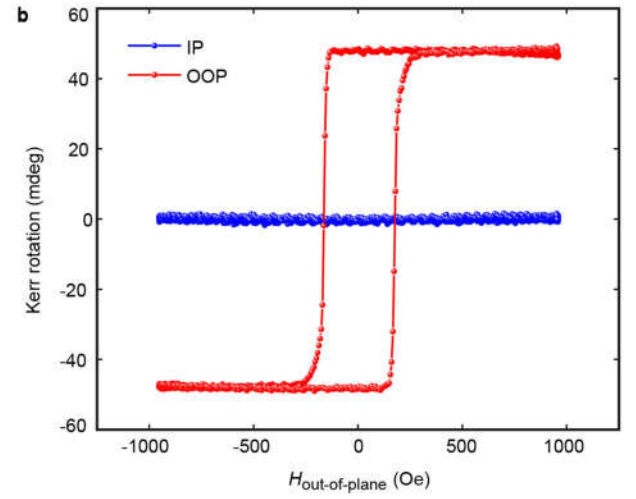
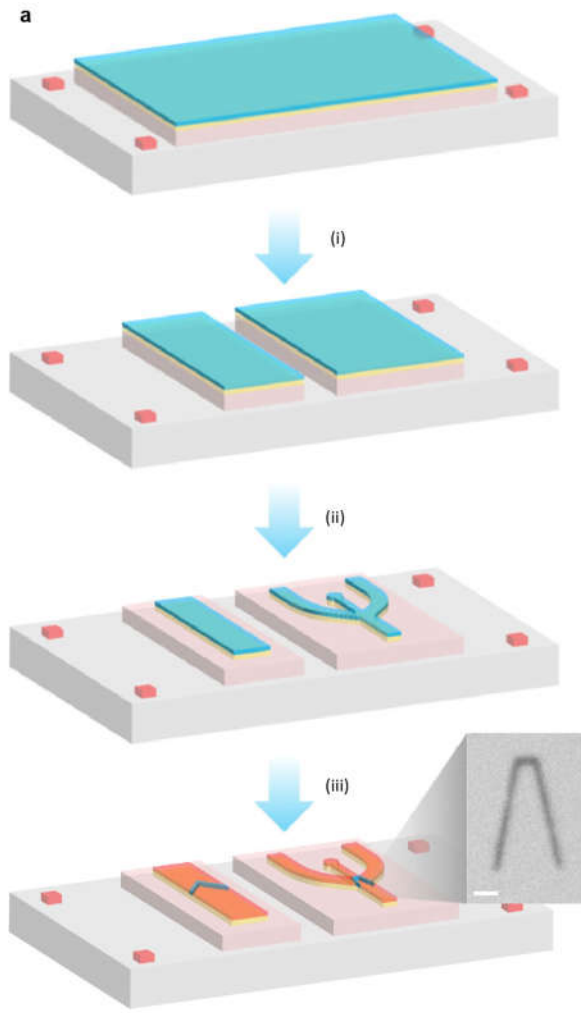
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2061-y>.

Correspondence and requests for materials should be addressed to Z.L., P.G. or L.J.H.

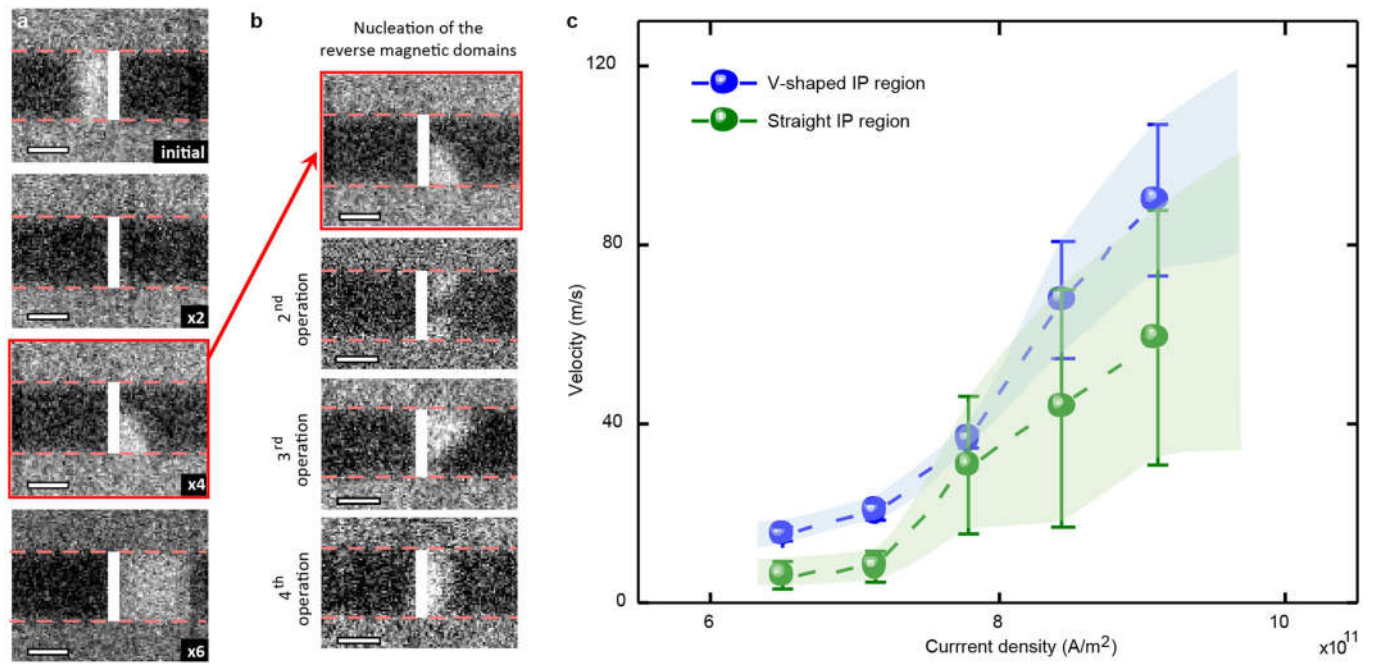
Peer review information Nature thanks See-Hun Yang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



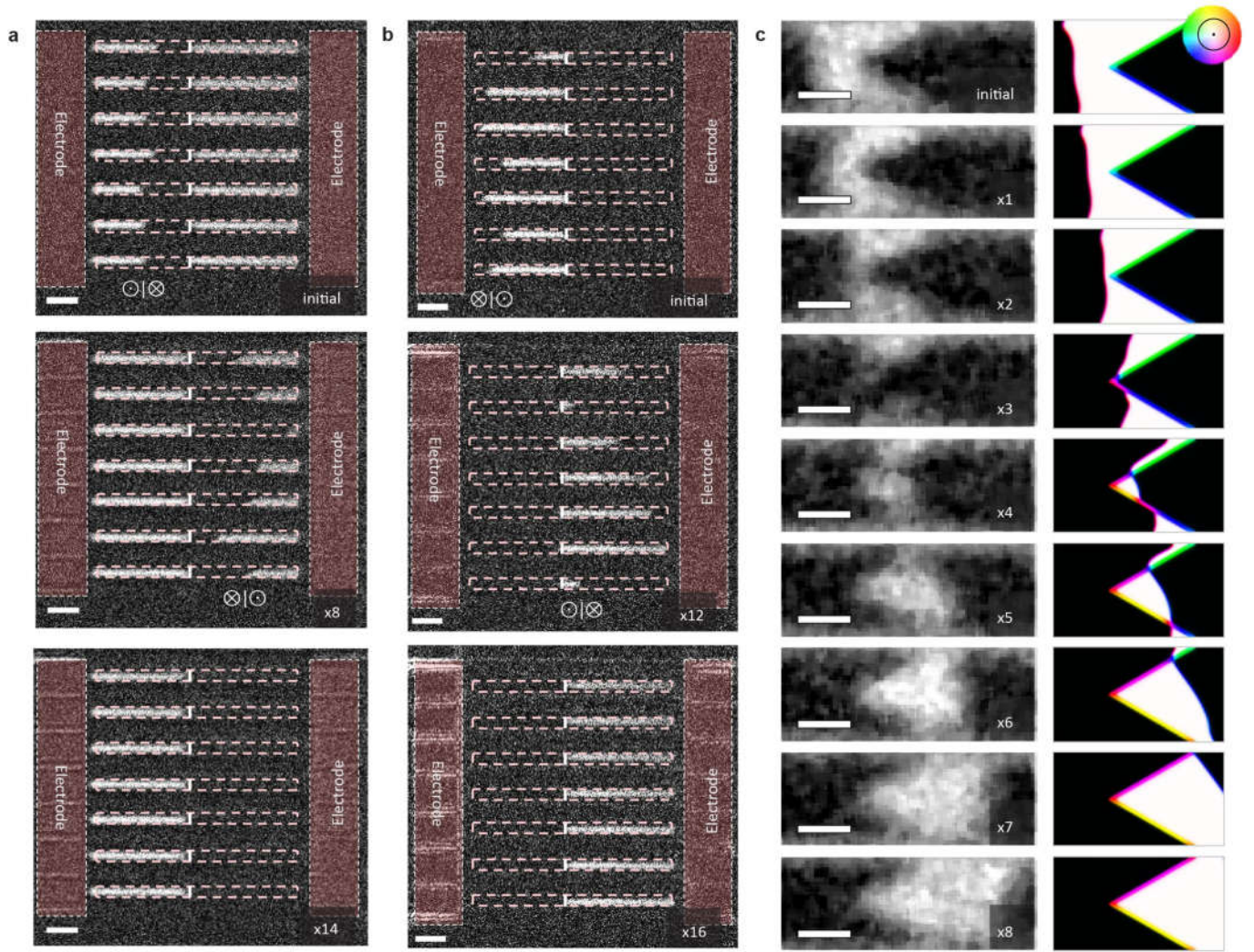
Extended Data Fig. 1 | Device fabrication and magnetic characterization. **a**, Schematics of main nanofabrication processes for magnetic DW logic circuits. (i) Ion milling of magnetic Pt/Co/Al multilayer to create magnetic strips, (ii) ion milling to produce magnetic racetracks and logic gates, and (iii) oxidation of the Al layer in the OOP regions. The inset shows an SEM image

of the 50-nm-wide PMMA mask used to protect the IP region of the NAND gate shown in Fig. 3a during oxygen plasma treatment. The scale bar is 100 nm. **b**, Polar MOKE measurement of the IP and OOP regions on application of an OOP magnetic field. **c**, Anomalous Hall measurement of the OOP region on application of an IP magnetic field.



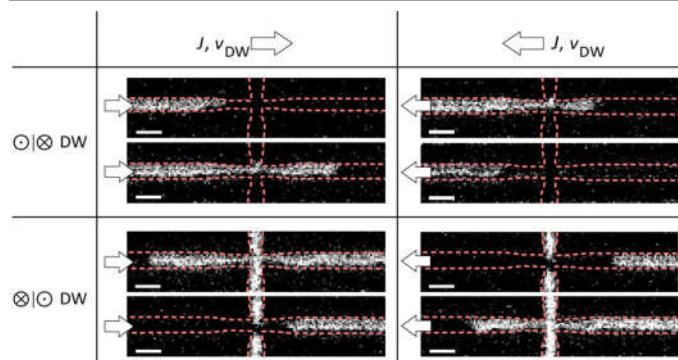
Extended Data Fig. 2 | DW inversion in a DW inverter with a straight IP region. **a**, STXM image sequence of DW inversion for an incident \otimes/\odot DW performed in a DW inverter with a straight IP region. Each XMCD image is captured after the application of two current pulses. The edges of the magnetic racetracks are indicated with red dashed lines and the positions of the inverters are indicated with solid white lines. The bright and dark regions in the XMCD images correspond to \odot and \otimes magnetization, respectively. The current

density and duration of the pulses are $1.1 \times 10^{12} \text{ A m}^{-2}$ and 1 ns, correspondingly. **b**, STXM images of the nucleation of reversed magnetic domains in the same DW inverter for four different operations. **c**, DW velocity in inverters with V-shaped and straight IP regions as a function of current density, determined from the experimental MOKE measurements. Error bars and shading represent the standard deviation of the DW velocity for 5 measurements. All the scale bars are 500 nm.

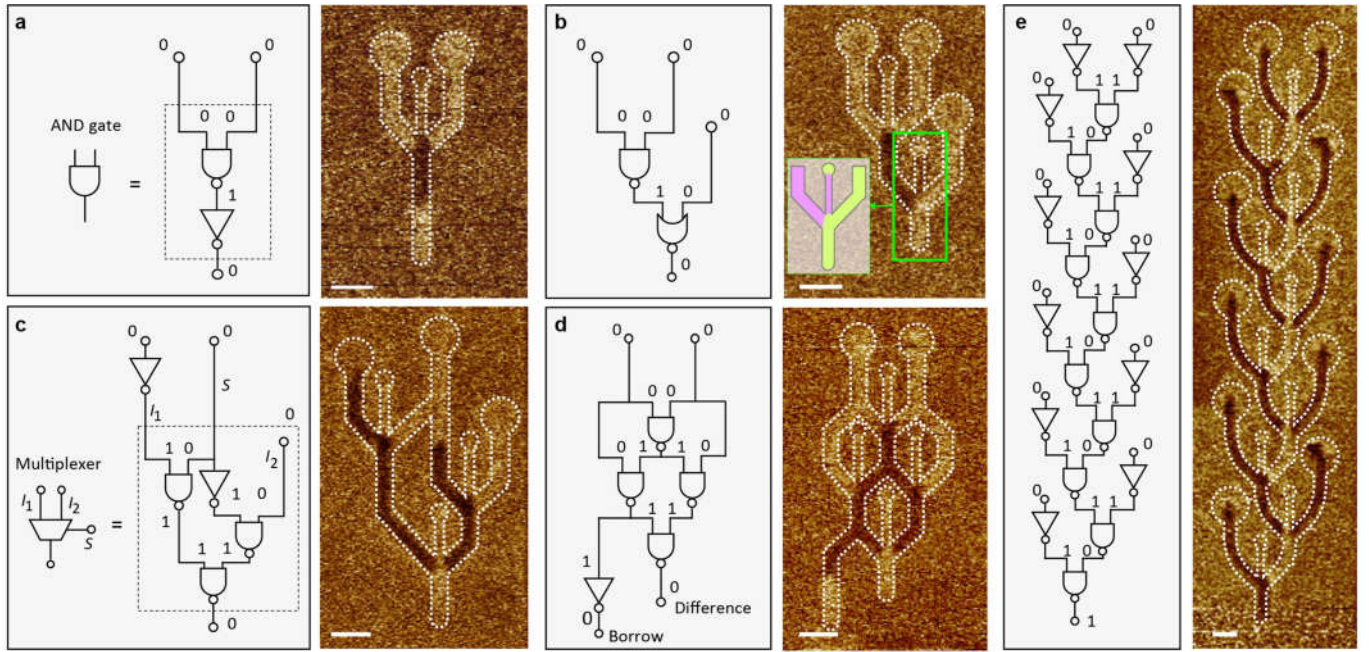


Extended Data Fig. 3 | Further experimental demonstration of the DW inverter. **a, b,** MOKE image sequences of DW inversion for incident $\odot|\otimes$ (**a**) and $\otimes|\odot$ (**b**) DWs performed in the same device used for Fig. 2c. The edges of the magnetic racetracks are indicated by red dashed lines and the positions of the inverters are indicated by white lines. The bright and dark regions in the magnetic racetracks in the MOKE images correspond to \odot and \otimes magnetization, respectively. The current density and pulse length of the applied current pulses are $7.5 \times 10^{11} \text{ A m}^{-2}$ and 50 ns, correspondingly. The number of applied current pulses is indicated. The entire image sequences are

shown in Supplementary Videos 1, 4. **c,** STXM image sequence of DW inversion for an incident $\otimes|\odot$ DW and corresponding micromagnetic simulation. Each XMCD image is captured after the application of one current pulse with a current density of $1.1 \times 10^{12} \text{ A m}^{-2}$ and a pulse length of 1 ns. The bright and dark regions in the XMCD images correspond to \odot and \otimes magnetization, respectively. For the simulated images, the IP directions of the magnetizations are given by the colour wheel, and the white and black regions correspond to \odot and \otimes magnetization, respectively. The scale bars in the MOKE images are $3 \mu\text{m}$ and those in the XMCD images with simulations are 500 nm.

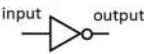
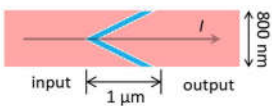
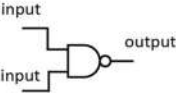
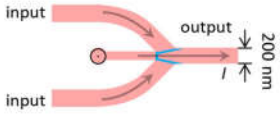
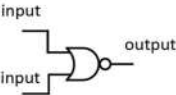
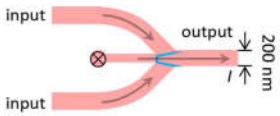
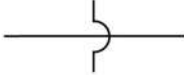

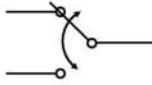
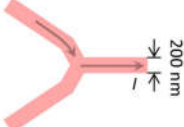
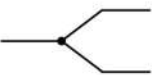
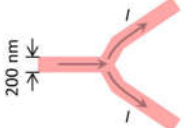


Extended Data Fig. 4 | Electrical control of DW motion through a cross structure. The directions of the current J and the DW motion are the same and are indicated for each image (current density $9 \times 10^{11} \text{ A m}^{-2}$, pulse length 30 ns). The edges of the cross structures are indicated with red dashed lines. The bright and dark regions in the cross structure in the MOKE images correspond to \odot and \otimes magnetization, respectively. All the scale bars are $1 \mu\text{m}$.

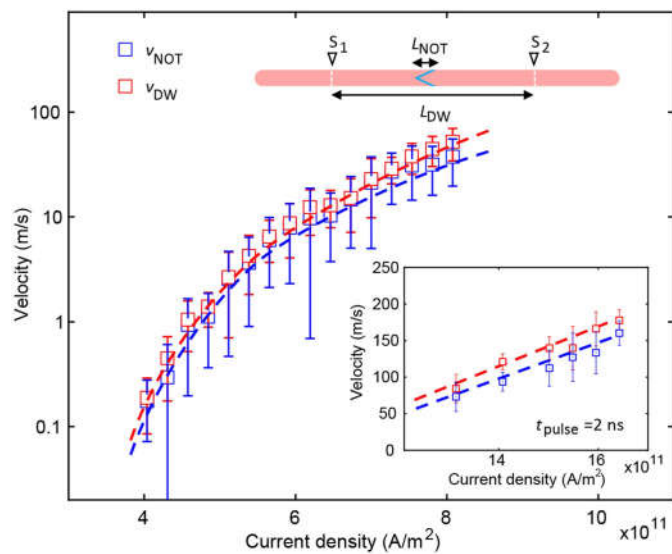


Extended Data Fig. 5 | Various cascaded DW logic circuits. a, AND gate fabricated by cascading one NAND gate and one NOT gate. **b**, Cascaded DW logic circuit with a NAND gate and a NOR gate. Green and purple in the schematic correspond to \odot and \otimes magnetization, respectively. There is an inverter placed in the bias reservoir of the NOR gate highlighted with the green box, giving a bias of '1', as shown in inset, whereas the bias for the NAND gate is '0'. **c**, Two-bit multiplexer constructed by cascading three NAND gates and

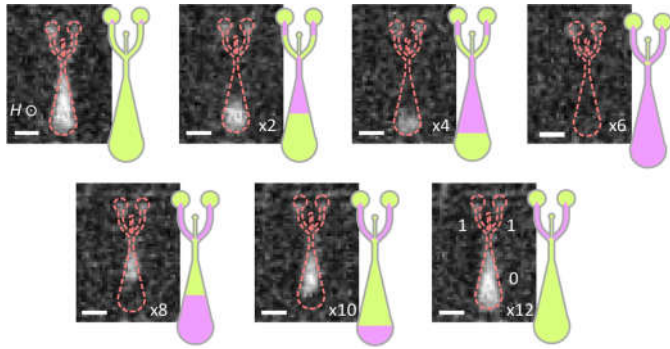
one NOT gate. **d**, Half-subtractor constructed by cascading four NAND gates and one NOT gate. **e**, Extensive cascaded DW logic circuits including 10 NAND gates and 11 NOT gates. The bright and dark areas in the device regions in the MFM images correspond to \odot and \otimes magnetization, respectively. The MFM images are captured after saturation with an OOP magnetic field to set the initial magnetization direction to \odot in all of the reservoirs, followed by current pulses to obtain the final states. All the scale bars are 500 nm.

Symbol	Magnetic DW logic
<div><p>NOT gate</p></div>	
<div><p>NAND gate</p></div>	
<div><p>NOR gate</p></div>	
<div><p>Cross-over</p></div>	
<div><p>Switch</p></div>	
<div><p>Fan-out</p></div>	

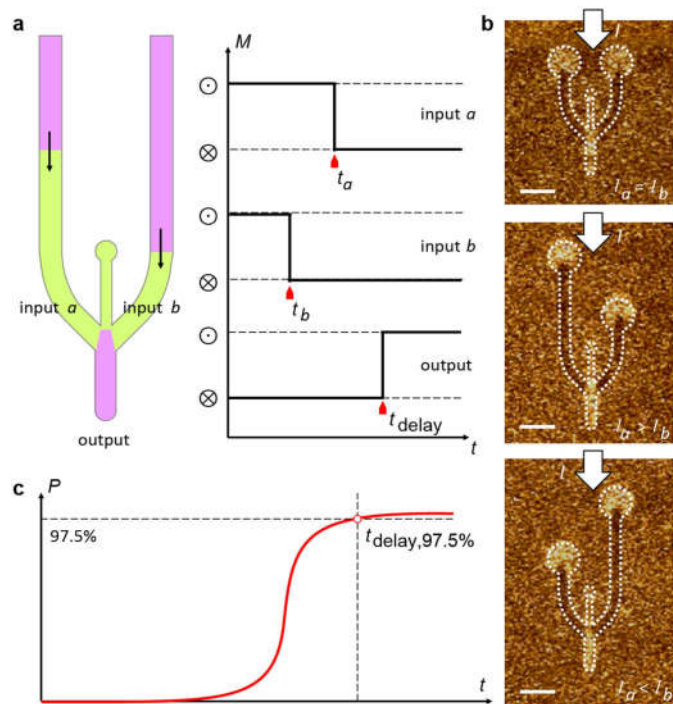
Extended Data Fig. 6 | Magnetic DW logic elements. Red and blue shaded regions indicate regions that have OOP and IP anisotropy, respectively. The direction of the current flow is indicated by black arrows. The dimensions of the magnetic DW logic elements used in the experiments are indicated.



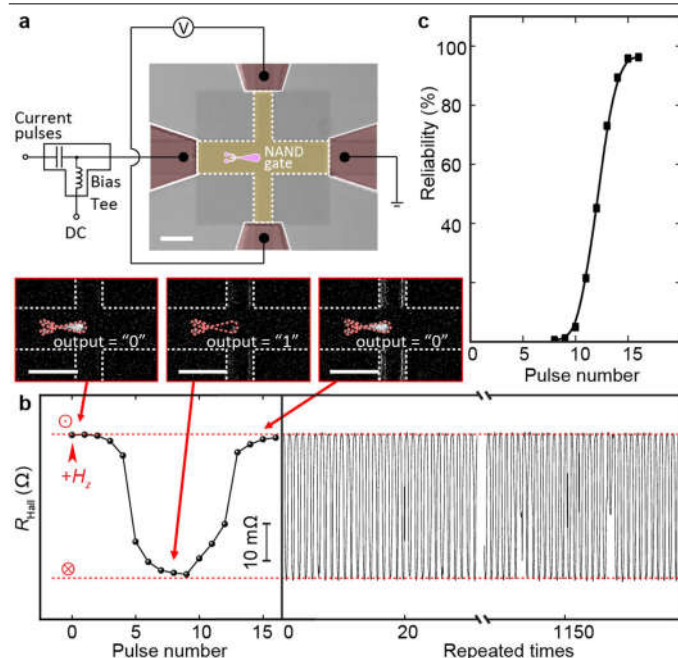
Extended Data Fig. 7 | DW velocity in a uniform OOP region of a racetrack and effective DW velocity in a NOT gate as a function of current density. Error bars represent the standard deviation of the DW velocity measured in 5 different devices.



Extended Data Fig. 8 | Time sequence of MOKE images of the NAND gate during operation and corresponding schematics. The NAND gate contains two inverters in each of the DW reservoirs and a bias set to '0'. The boundaries of the NAND gate are indicated by red dashed lines. The two V-shaped inverters in the DW reservoirs are associated with small reversed domains in the initial state (purple triangles) resulting from the chiral coupling. A sequence of MOKE images is captured and each image is taken following two current pulses with a current density of $7.5 \times 10^{11} \text{ A m}^{-2}$ and a pulse length of 30 ns. The bright and dark regions in the gate structure in the MOKE images correspond to \odot and \otimes magnetization, respectively. In the schematics, green and purple correspond to \odot and \otimes magnetization, respectively. The two DW reservoirs and the bias are set to logical value '0' by applying an OOP magnetic field of 1 kOe. All the scale bars are $1 \mu\text{m}$.



Extended Data Fig. 9 | Propagation delay time for DW logic. **a**, Schematic showing the use of the propagation delay time to improve the operational reliability of logic operation in a NAND gate. Green and purple correspond to \odot and \otimes magnetization, respectively. **b**, MFM images of NAND gates with different input racetrack lengths (top, $l_a = l_b$; middle, $l_a > l_b$; bottom, $l_a < l_b$; where l_a and l_b represent the racetrack lengths of input a and b , respectively). The bright and dark regions correspond to \odot and \otimes magnetization, respectively. The MFM images are captured after saturation with an OOP magnetic field to set the initial magnetization direction to \odot in all of the reservoirs, followed by application of current pulses to obtain the final states. The direction of the current flow is indicated. **c**, Schematic showing the dependence of the probability of giving a correct output as a function of propagation delay time. All the scale bars in the MFM images are 500 nm.



Extended Data Fig. 10 | Hall measurement of logic operation in a NAND gate.

a, Schematic and optical microscope image of the device. Red, yellow and purple colours in the image represent the regions with electrodes, the Pt cross and the NAND gate, respectively. **b**, Hall resistance as a function of pulse number and corresponding MOKE images. Left, typical evolution of the Hall resistance with increasing number of pulses. Right, the first 30 and last 30 of the 1,172 repeated measurements. The Hall resistance levels for \odot and \otimes output magnetizations are indicated by the red dashed lines. An OOP magnetic field is applied to set the initial state at the beginning of each measurement, indicated with the red arrow. The bright and dark regions in the gate structure in the MOKE images correspond to \odot and \otimes magnetization, respectively. The boundaries of the NAND gate are indicated by red dashed lines. The current density and pulse length of the applied current pulses are $7.5 \times 10^{11} \text{ A m}^{-2}$ and 30 ns, respectively. **c**, Operational reliability as a function of the number of current pulses. All the scale bars are $2 \mu\text{m}$.

Wafer-scale single-crystal hexagonal boron nitride monolayers on Cu (111)

<https://doi.org/10.1038/s41586-020-2009-2>

Received: 23 June 2019

Accepted: 10 December 2019

Published online: 4 March 2020

 Check for updates

Tse-An Chen^{1,10}, Chih-Piao Chuu^{1,10}, Chien-Chih Tseng², Chao-Kai Wen², H.-S. Philip Wong¹, Shuangyuan Pan³, Rongtan Li^{4,5}, Tzu-Ang Chao^{1,2}, Wei-Chen Chueh², Yanfeng Zhang³, Qiang Fu⁴, Boris I. Yakobson^{6,7,8}, Wen-Hao Chang^{2,9} & Lain-Jong Li^{1,8}

Ultrathin two-dimensional (2D) semiconducting layered materials offer great potential for extending Moore's law of the number of transistors in an integrated circuit¹. One key challenge with 2D semiconductors is to avoid the formation of charge scattering and trap sites from adjacent dielectrics. An insulating van der Waals layer of hexagonal boron nitride (hBN) provides an excellent interface dielectric, efficiently reducing charge scattering^{2,3}. Recent studies have shown the growth of single-crystal hBN films on molten gold surfaces⁴ or bulk copper foils⁵. However, the use of molten gold is not favoured by industry, owing to its high cost, cross-contamination and potential issues of process control and scalability. Copper foils might be suitable for roll-to-roll processes, but are unlikely to be compatible with advanced microelectronic fabrication on wafers. Thus, a reliable way of growing single-crystal hBN films directly on wafers would contribute to the broad adoption of 2D layered materials in industry. Previous attempts to grow hBN monolayers on Cu (111) metals have failed to achieve mono-orientation, resulting in unwanted grain boundaries when the layers merge into films^{6,7}. Growing single-crystal hBN on such high-symmetry surface planes as Cu (111)^{5,8} is widely believed to be impossible, even in theory. Nonetheless, here we report the successful epitaxial growth of single-crystal hBN monolayers on a Cu (111) thin film across a two-inch *c*-plane sapphire wafer. This surprising result is corroborated by our first-principles calculations, suggesting that the epitaxial growth is enhanced by lateral docking of hBN to Cu (111) steps, ensuring the mono-orientation of hBN monolayers. The obtained single-crystal hBN, incorporated as an interface layer between molybdenum disulfide and hafnium dioxide in a bottom-gate configuration, enhanced the electrical performance of transistors. This reliable approach to producing wafer-scale single-crystal hBN paves the way to future 2D electronics.

First, a single-crystal Cu (111) thin film on a wafer is needed. Single-crystal Cu in thick foils can be achieved through recrystallization induced by implanted seeds^{5,9}. However, for the formation of Cu (111) thin film on a wafer, the crystallinity relies strongly on the underlying substrate lattices. Here we used a *c*-plane sapphire as the substrate, on which we sputtered a 500-nm-thick polycrystalline Cu film and then carried out extensive thermal annealing to achieve single-crystal Cu (111) films¹⁰. One challenge is that Cu (111) tends to form twin grains separated by twin-grain boundaries, through kinetic growth processes. Figure 1a illustrates the atomic arrangements for the typical twinned Cu (111) structure. We find that post-annealing at a high temperature (1,040–1,070 °C) in the presence of hydrogen is the key to removing the twin grains, consistent with

recent reports^{10,11}. Figure 1b, c shows optical micrographs and electron backscatter diffraction (EBSD) patterns for the Cu (111) thin films after annealing at 1,000 °C and 1,050 °C. The EBSD results (see also Extended Data Fig. 1a, b) confirm the coexistence of twinned Cu (111) polycrystals with 0° and 60° in-plane misorientation for the Cu thin films annealed at 1,000 °C. The in-plane misorientation was removed after annealing at 1,050 °C, producing single-crystal Cu (111). The X-ray diffraction results (Extended Data Fig. 1c–f) also consistently illustrate our success in obtaining single-crystal Cu (111) thin films. Note that the Cu (111) is preferentially formed with a thinner Cu film, but a sufficiently thick Cu film is necessary to prevent Cu evaporation during subsequent hBN growth. Thus, there is an optimal Cu thickness (around 500 nm) for single-crystal hBN growth.

¹Corporate Research, Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, Taiwan. ²Department of Electrophysics, National Chiao Tung University, Hsinchu, Taiwan. ³Department of Materials Science and Engineering, College of Engineering, Peking University, Beijing, China. ⁴State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China. ⁵University of Chinese Academy of Sciences, Beijing, China. ⁶Department of Materials Science and Nanoengineering, Rice University, Houston, TX, USA. ⁷Department of Chemistry, Rice University, Houston, TX, USA. ⁸Smalley-Curl Institute for Nanoscale Science and Technology, Rice University, Houston, TX, USA. ⁹Center for Emergent Functional Matter Science (CEFMS), National Chiao Tung University, Hsinchu, Taiwan. ¹⁰These authors contributed equally: Tse-An Chen, Chih-Piao Chuu. [✉]e-mail: biy@rice.edu; whchang@mail.nctu.edu.tw; ljli@tsmc.com

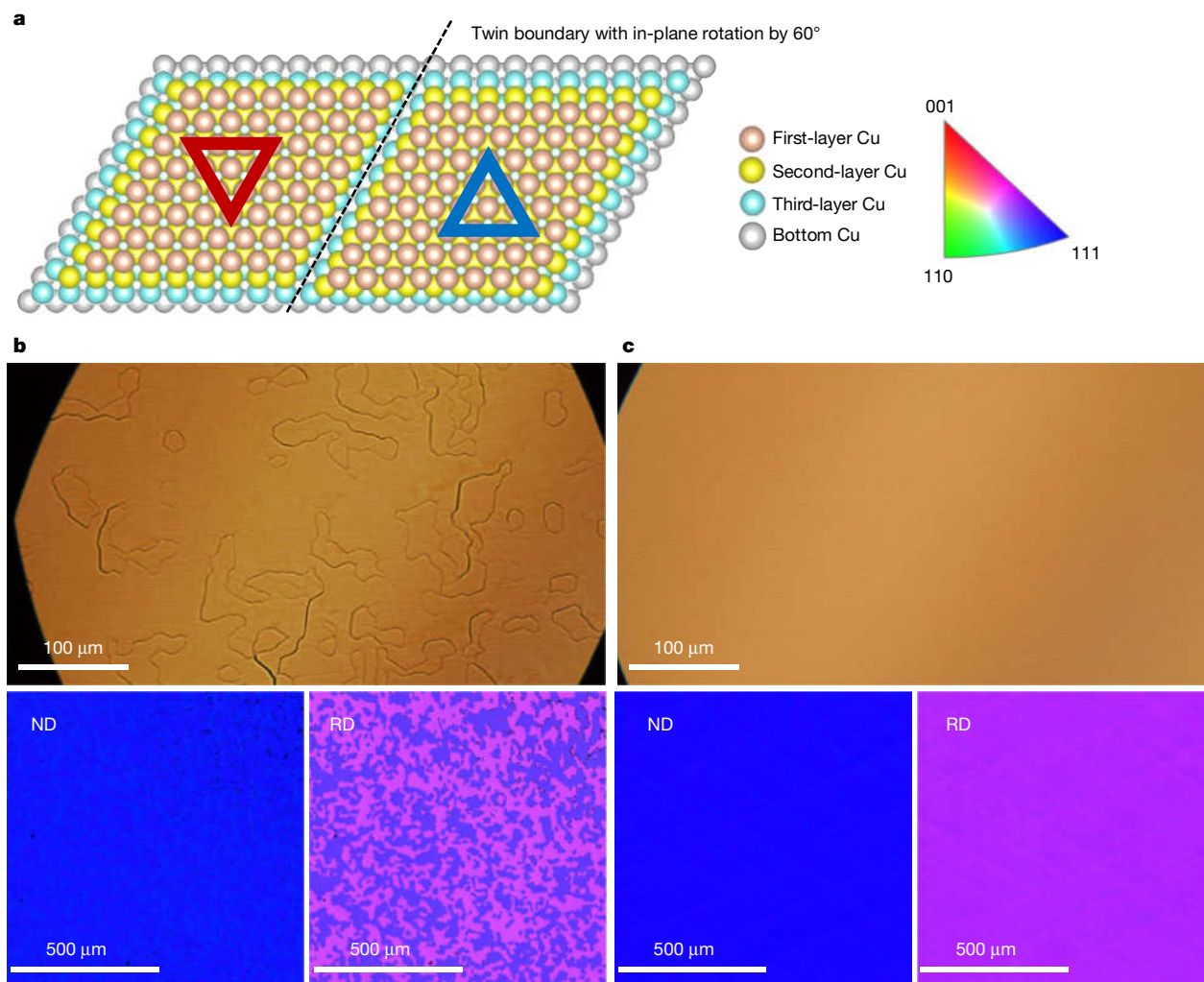


Fig. 1 | Cu (111) lattice orientation on c-sapphire substrates. **a**, Diagram showing twinned Cu (111) (top view) with an in-plane rotation of 60°; the blue and red triangles enclose areas with the same Cu stacking configuration in adjacent twinned grains. The right-hand coloured triangle is the legend for the inverse pole figure (IPF) maps at the bottom of panels **b**, **c**. **b**, Optical microscope image (top) and EBSD IPF mappings ($1 \times 1 \text{ mm}^2$; bottom) of a Cu thin

film annealed at 1,000 °C for 1 h, where the normal direction (ND) and rolling direction (RD) indicate the presence of twinned Cu (111) polycrystals with in-plane misorientation by 60°. **c**, If the Cu thin film is annealed at 1,050 °C for 1 h, the in-plane orientation is unified, showing the formation of single-crystal Cu (111) without any twinned grains.

Achieving the growth of mono-oriented hBN triangular flakes is an essential step towards obtaining wafer-scale single-crystal hBN. Owing to the sixfold symmetry of Cu (111), the van der Waals registry of hBN to Cu (111) leads to two sets of energy-minimal configurations (whose orientation differs by 60° or 180°, an inversion) with almost degenerate binding energies. Thus it is commonly believed that confining hBN flakes to mono-orientation on such a high-symmetry surface is impossible^{5,8}. Our experiments reveal that the energy degeneracy can be lifted in the presence of spontaneously present top-layer Cu step edges. The growth of an hBN monolayer is carried out by flowing ammonia borane precursors onto the one-inch single-crystal Cu (111) thin film/sapphire in a hot-wall chemical vapour deposition (CVD) furnace. An optical micrograph of monolayer hBN triangular flakes grown on a Cu (111) thin film with twin grains (Fig. 2a) shows that hBN flakes orient to the same direction on one twin and to the opposite direction (or 60° in-plane rotation along the *z* axis) on the counterpart twin (Extended Data Fig. 2). Figure 2b displays optical micrographs of hBN flakes grown on a single-crystal Cu (111) thin film without twin grains, where almost all of the triangles are unidirectionally aligned (see also Extended Data Fig. 3 for statistical analysis of the orientation distribution of hBN triangular flakes). The observation of mono-orientation on an individual

single-crystal Cu (111) grain clearly indicates the existence of an energy-minimized hBN–Cu (111) configuration. Therefore, eliminating the twin grains in Cu (111) will ensure the growth of single-crystal hBN on it.

To verify the single crystallinity, we characterized the hBN monolayer merged from mono-oriented triangles by microspot low-energy electron diffraction (μ -LEED), using a probe size of around 3 μm at 80 sites across the one-inch wafer. Figure 2c displays the μ -LEED patterns from nine randomly selected sites. All results reveal an hBN monolayer that is unidirectionally aligned with the Cu (111) surfaces, indicating that their single crystallinity strictly follows the Cu (111) lattices. The atomically resolved scanning tunnelling microscopy (STM) image of hBN on Cu (111) in Fig. 2d shows a perfect hBN lattice with a measured lattice constant of $2.50 \pm 0.1 \text{ Å}$, consistent with the theoretical value of 2.5 Å. We probed more than 20 locations, and all STM images show the same hBN lattice orientation (Extended Data Fig. 4). We did not observe any grain boundaries formed by adjacent misoriented hBN domains, suggesting the single-crystalline nature of hBN. We note that, in some areas, moiré patterns arise from the lattice mismatch and/or relatively small rotation (within 1.5°) between hBN and the underlying Cu (111) substrate (Extended Data Fig. 5a–f). The magnified atomic-resolution image at the moiré boundary areas reveals that the hBN

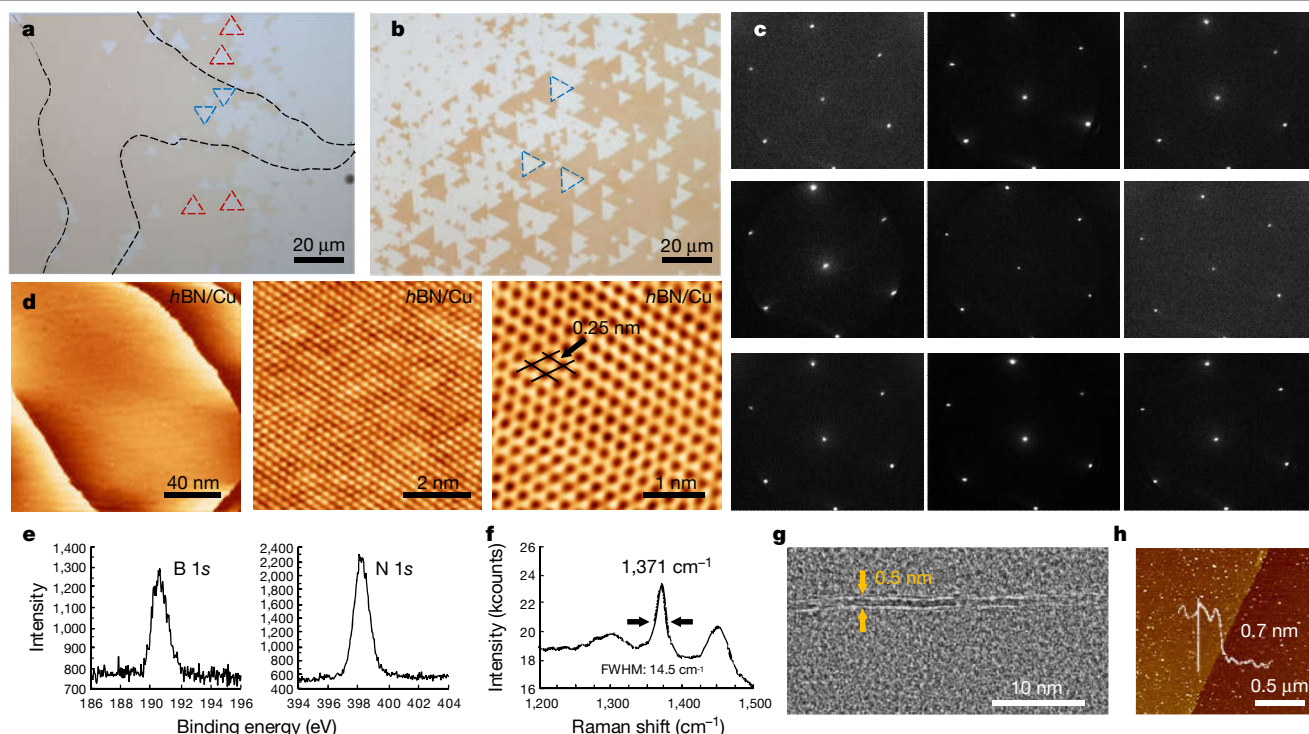


Fig. 2 | Growth and atomic structures of single-crystal hBN on Cu (111) films.

a, Optical microscope image of hBN grown on different Cu (111) grains. Black dashed lines indicate twin grain boundaries. Oppositely oriented flakes are marked by red and blue dashed triangles. **b**, Mono-oriented hBN flakes on single-crystal Cu (111) films. **c**, μ -LEED patterns of hBN monolayers at nine different areas randomly selected from the $1.5 \times 1.5 \text{ cm}^2$ sample surface. All μ -LEED patterns show that the hBN monolayers have the same orientation as the Cu (111) surfaces. **d**, Left to right, large-scale STM image ($V_{\text{tip}} = -1.008 \text{ V}$; $I_{\text{tip}} = 3.90 \text{ nA}$; $T = 300 \text{ K}$) and atomic-scale STM images ($V_{\text{tip}} = -0.003 \text{ V}$; $I_{\text{tip}} = 54.508 \text{ nA}$; $V_{\text{tip}} = -0.003 \text{ V}$, $I_{\text{tip}} = 46.50 \text{ nA}$; $T = 300 \text{ K}$) of hBN/Cu (111) with

measured lattice constant $2.50 \pm 0.1 \text{ \AA}$. We measured the height of each step from the left-hand image to be roughly 2 \AA . **e**, XPS spectra measured from an as-grown monolayer hBN film on Cu (111)/sapphire. The binding energies of the B 1s and N 1s orbitals at 190.4 eV and 398.0 eV confirm the formation of hBN. **f**, Raman spectrum of a transferred hBN film, where the E_{2g} mode at $1,371 \text{ cm}^{-1}$ with a full-width at half-maximum (FWHM) of 14.5 cm^{-1} confirms that the hBN is a monolayer. **g**, Cross-sectional TEM image of a monolayer hBN transferred onto SiO_2/Si , where the thickness of hBN is around 0.5 nm . **h**, AFM image of a single-crystal hBN film transferred onto a SiO_2/Si substrate.

presents perfect lattice coherence at the patching boundary (Extended Data Fig. 5g–j), indicating that the formation of moiré patterns does not affect the overall hBN orientation. We believe that the hBN completes the single-crystal growth at high temperatures, and that the strain associated with sample cooling after growth results in the formation of local moiré pattern. Other characterizations, including X-ray photoelectron spectroscopy (XPS) and Raman spectroscopy, prove the B–N chemical bonding structures (Fig. 2e, f). Transmission electron microscope (TEM) and atomic force microscope (AFM) images consistently show that the as-grown hBN is indeed a monolayer (Fig. 2g, h).

We recognize that, once the Cu (111) thin films are prepared and formed untwinned at $1,050 \text{ }^\circ\text{C}$ as described above, the mono-oriented growth of hBN flakes can be realized at various growth temperatures ranging from $995 \text{ }^\circ\text{C}$ to $1,070 \text{ }^\circ\text{C}$ (Extended Data Fig. 6). However, a lower growth temperature ($995 \text{ }^\circ\text{C}$ to $1,010 \text{ }^\circ\text{C}$) usually leads to lower-quality hBN flakes, which are easily oxidized in a subsequent oxidation test at $150 \text{ }^\circ\text{C}$ in air. Therefore, we used a higher growth temperature (typically $1,050 \text{ }^\circ\text{C}$) to ensure high-quality single-crystal hBN growth.

To explain the preferred orientation of hBN on Cu (111), we consider a small and rigid B_6N_7 molecule (that is, an energetically favourable N-terminated three-ring structure¹²) as the probe seed. We first examine the effect of plane-to-plane epitaxy, using density functional theory (DFT) to calculate the binding energies of six typical atomic stacking configurations (Fig. 3a), where N_iB_{iii} , $\text{N}_{iii}\text{B}_{ii}$ and N_{ii}B_i are defined as the 0° orientation, and N_iB_{ii} , $\text{N}_{ii}\text{B}_{iii}$ and N_{iii}B_i are the 60° (inverted) orientation. The notation N_iB_j represents the stacking of N atoms in registry with (above) the Cu atoms in the i th layer, while B atoms register with the

Cu atoms in the j th layer. The calculations show that the stacking with N atoms on top of the first-layer Cu atoms (N_iB_{iii} (0°) and N_iB_{ii} (60°)) exhibits the lowest energy, while B atoms on top of the first-layer Cu atoms (N_{ii}B_i (0°) and N_{iii}B_i (60°)) are energetically unfavourable. The preferential registrations reflect the electron affinity of the B and N atoms, which leads to attractive (or repulsive) Coulomb interactions between N (or B) atoms and the first-layer Cu atoms, and hence affects the structural stability. We find that the lowest-energy structures for 0° (N_iB_{iii}) and 60° (N_iB_{ii}) orientations exhibit an energy difference of only 0.05 eV or so, much smaller than the thermal energy $k_B T$ at the growth temperature (roughly 0.1 eV), indicating that the plane-to-plane registry is insufficient to achieve mono-oriented growth, in agreement with simulations¹³.

In fact the Cu (111) surface is not perfectly flat and many terraced meandering steps exist, as revealed in STM images (Fig. 2d and Extended Data Fig. 7a–d). Recent theory showed that one must consider the role of these step edges in guiding hBN growth¹⁴. Other work⁵ suggests that docking at the vicinal step edges on the Cu (110) surface governs the single-crystal hBN growth, based on the assumption that the Cu terrace steps trend only up or down all the way across the whole vicinal surface of the Cu foil. However, our STM results clearly reveal that the terrace steps of the Cu (111) surface trend both up and down across the wafer, and that the edge-docking can seemingly yield hBN in both directions, unless the binding energies differ enough to favour one direction over another. To capture this in our model, we add an extra layer of Cu atoms (red in Fig. 3) on top of the first layer, forming two opposite step edges (A- and B-step edges in Fig. 3a and Extended Data Fig. 7f). This restricts

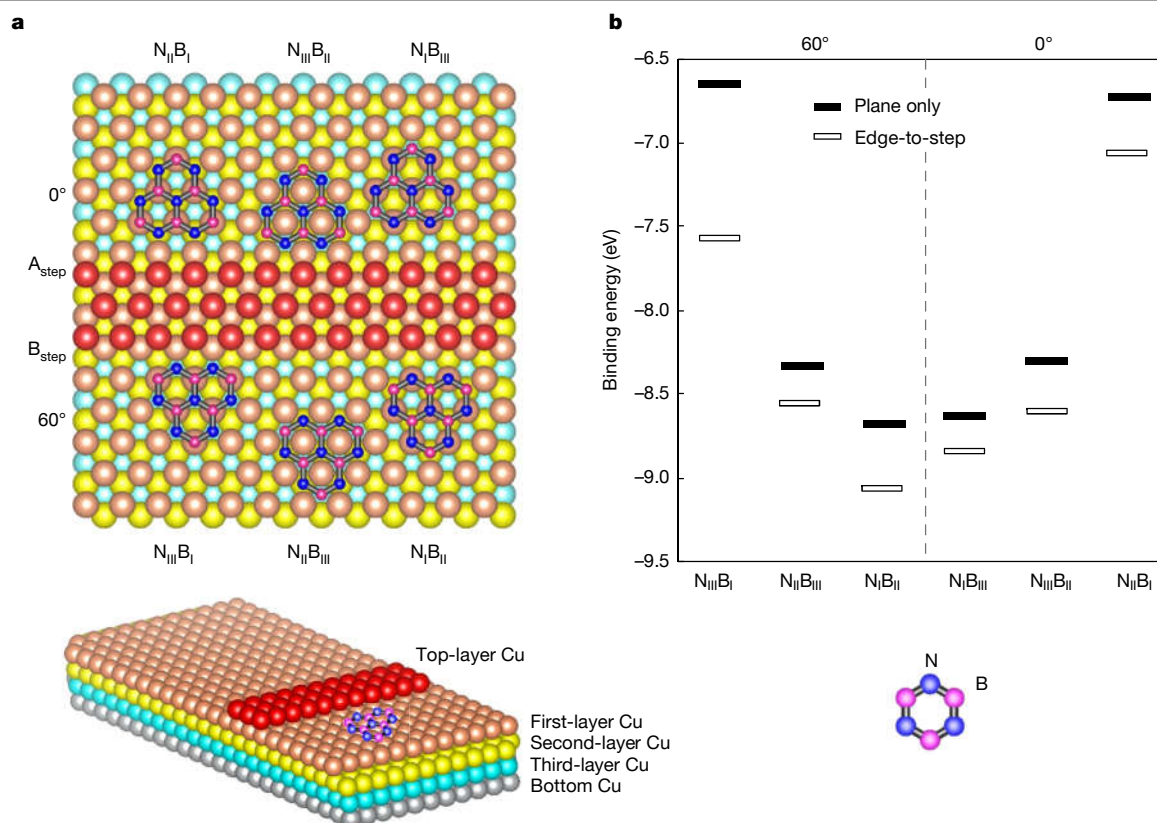


Fig. 3 | DFT calculations of epitaxy at nucleation, with and without considering step-edge docking. a, Lowest-energy atomic arrangements for six B_6N_7 -Cu(111) configurations, considering the edge docking to two typical step-edge terminations (A and B) of top-layer Cu(111). **b**, Calculated binding

energies for the six B_6N_7 -Cu(111) configurations, with and without including the edge-to-step epitaxy effect. The bottom image is a three-dimensional view of the $N_I B_{II}$ -Cu(111) configuration.

the B_6N_7 seed to the 0° (or 60°) orientation when docking to the A (or B) steps (Fig. 3a). In Fig. 3a, b the distance between the Cu step edge and the B_6N_7 zigzag edge for each configuration has been determined by energy minimization (Extended Data Fig. 8). The resulting binding energy for each configuration changes in a subtle yet very important

way with the presence of Cu step edges (Fig. 3b): the two configurations $N_I B_{II}$ (60°) and $N_I B_{III}$ (0°), which were nearly degenerate when considering only plane-to-plane epitaxy, are now separated by δE values of approximately 0.23 eV, this value raising in proportion with the docking length, rapidly amplifying the Boltzmann selectivity factor

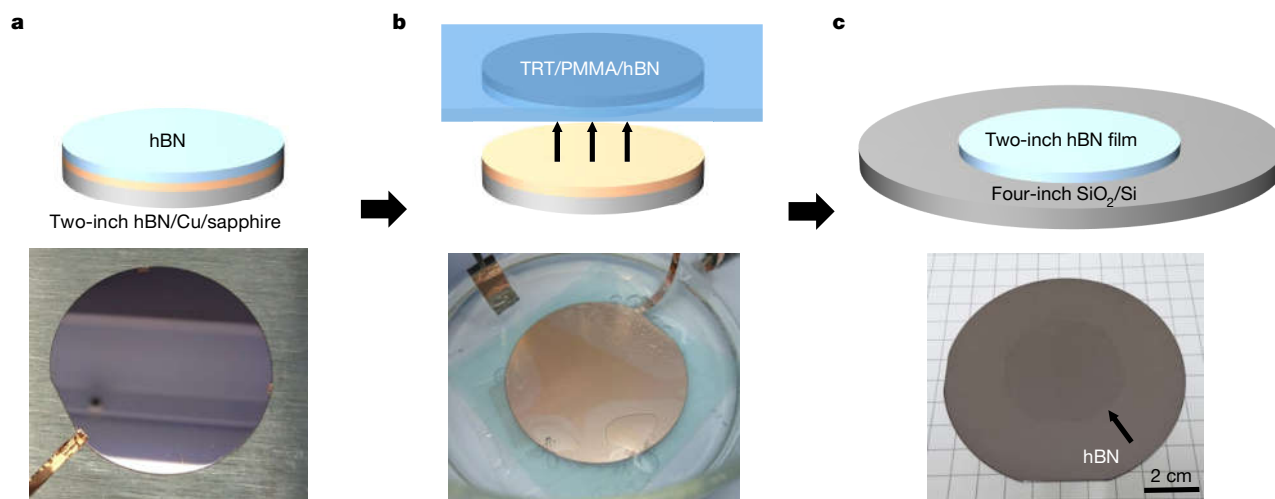


Fig. 4 | Schematic and photographs illustrating the wafer-scale hBN transfer processes. a, As-grown two-inch hBN film on a Cu(111)/sapphire wafer. Top, schematic; bottom, photograph. **b**, Electrochemical delamination, using an aqueous solution of NaOH (1M) as the electrolyte, the Cu layer in the TRT/PMMA/hBN/Cu(111)/sapphire stack as the cathode, and a platinum (Pt) foil as

the anode, with an applied DC voltage of 4 V. During this process, the TRT/PMMA/hBN stacked film is detached from the Cu(111)/sapphire through the generation of hydrogen bubbles at the hBN/Cu interface. PMMA, poly(methyl methacrylate); TRT, thermal release tape. **c**, Two-inch monolayer hBN film transferred onto a four-inch SiO_2/Si wafer.

$e^{\delta E/k_B T}$ (for a contact length of just five to six hexagons, the Boltzmann selectivity factor increases to more than 10^3 ; Extended Data Fig. 9). Such an energy difference apparently ensures the mono-orientated growth. Our STM results (Extended Data Fig. 7) show that all meandering steps are rather curved and locally rugged, so that they all consist of segments of A and B types. BN seeds should kinetically nucleate at the A to B corner while docking to stronger binding sites, B types, with proper orientation (Extended Data Fig. 7e). The simulations, together with experimental results, indicate that Cu (111) surfaces with step edges are the key to achieving single-crystal hBN growth.

Following our success in growing single-crystal hBN on a one-inch Cu (111) thin film, we further scaled the growth to a two-inch wafer, as depicted in Fig. 4a. Given that the interaction between a fully grown hBN layer and Cu (111) is limited to weak van der Waals forces, the detachment of wafer-scale hBN can be achieved by polymer-assisted transfer with the help of electrochemical processes^{15,16} (Fig. 4b and Extended Data Fig. 10). Figure 4c shows a photograph of a two-inch hBN monolayer transferred onto a four-inch SiO₂/Si wafer. We show that the growth of single-crystal wafer-scale hBN on Cu (111) thin films is scalable and much more cost-effective than using thick Cu foils or other metals, and thus could be a preferred approach for the microelectronics industry. The availability of wafer-scale single-crystal hBN should stimulate and enable further research and development of futuristic 2D electronics. We constructed monolayer MoS₂ field-effect transistors (FETs) with and without single-crystal and polycrystalline hBN as an interface dielectric in a bottom-gate configuration (Extended Data Fig. 11). The enhancement of mobility in MoS₂ and the suppression of hysteresis are substantial in the device with a single-crystal hBN monolayer, suggesting its promise for 2D-based transistors.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2009-2>.

1. Li, M.-Y., Su, S.-K., Wong, H.-S. P. & Li, L.-J. How 2D semiconductors could extend Moore's law. *Nature* **567**, 169–170 (2019).
2. Novoselov, K. S., Mishchenko, A., Carvalho, A. & Neto, A. H. C. 2D materials and van der Waals heterostructures. *Science* **353**, aac9439 (2016).
3. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
4. Lee, J. S. et al. Wafer-scale single-crystal hexagonal boron nitride film via self-collimated grain formation. *Science* **362**, 817–821 (2018).
5. Wang, L. et al. Epitaxial growth of a 100-square-centimetre single-crystal hexagonal boron nitride monolayer on copper. *Nature* **570**, 91–95 (2019).
6. Uchida, Y., Iwazako, T., Mizuno, S., Tsuji, M. & Ago, H. Epitaxial chemical vapour deposition growth of monolayer hexagonal boron nitride on a Cu (111)/sapphire substrate. *Phys. Chem. Chem. Phys.* **19**, 8230–8235 (2017).
7. Song, X. et al. Chemical vapor deposition growth of large-scale hexagonal boron nitride with controllable orientation. *Nano Res.* **8**, 3164–3176 (2015).
8. Li, J. et al. Growth of polar hexagonal boron nitride monolayer on nonpolar copper with unique orientation. *Small* **12**, 3645–3650 (2016).
9. Jin, S. et al. Colossal grain growth yields single-crystal metal foils by contact-free annealing. *Science* **362**, 1021–1025 (2018).
10. Verguts, K. et al. Epitaxial Al₂O₃ (0001)/Cu (111) template development for CVD graphene growth. *J. Phys. Chem. C* **120**, 297–304 (2016).
11. Deng, B. et al. Wrinkle-free single-crystal graphene wafer grown on strain-engineered substrates. *ACS Nano* **11**, 12337–12345 (2017).
12. Liu, Y., Bhowmick, S. & Yakobson, B. I. BN white graphene with “colorful” edges: the energies and morphology. *Nano Lett.* **11**, 3113–3116 (2011).
13. Zhao, R., Zhao, X., Liu, Z., Ding, F. & Liu, Z. Controlling the orientations of hBN during growth on transition metals by chemical vapor deposition. *Nanoscale* **9**, 3561–3567 (2017).
14. Bets, K. V., Gupta, N. & Yakobson, B. I. How the complementarity at vicinal steps enables growth of 2D monocrystals. *Nano Lett.* **19**, 2027–2031 (2019).
15. Gao, L. et al. Repeated growth and bubbling transfer of graphene with millimetre-size single-crystal grains using platinum. *Nat. Commun.* **3**, 699 (2012).
16. Kim, G. et al. Growth of high-crystalline, single-layer hexagonal boron nitride on recyclable platinum foil. *Nano Lett.* **13**, 1834–1839 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Growth of Cu (111) thin films

The as-received sapphire substrates were first etched in a mixed $\text{H}_2\text{SO}_4/\text{H}_3\text{PO}_4$ aqueous solution at 300 °C for 20 min. The etched sapphire substrates were cleaned by immersion in ultra-pure water for 5 min. After cleaning, the sapphire substrates were loaded into a sputtering chamber for Cu deposition. The chamber was maintained at room temperature under an argon pressure of 0.3 mtorr, giving rise to a Cu deposition rate of 2 nm s^{-1} (ref. ¹⁰).

Chemical vapour deposition of hBN

The monolayer hBN films were grown in a three-inch furnace tube with three heating zones using low-pressure chemical vapour deposition (LPCVD). The two-inch Cu (111)/sapphire substrate was placed in the central heating zone of the main chamber. Ammonia borane (97%, roughly 60 mg) was used as the precursor and loaded into a subchamber at the upstream side of the main chamber. The furnace was first pumped down to a base pressure of 5.0 torr. Before growth, the substrate was annealed at 1,050 °C for 60 min under a hydrogen gas flow of 300 standard cubic centimetres per minute (sccm). The subchamber (with precursors) was heated to 85 °C using a heating belt and maintained there for 30 min. The precursor was then introduced into the main chamber for 30 min in order to grow hBN on the Cu (111)/sapphire substrates, with the substrates facing downwards. After hBN growth, the subchamber was closed, and the main chamber was naturally cooled to room temperature under a hydrogen gas flow of 30 sccm.

Transfer of hBN films onto arbitrary substrates

The as-grown monolayer hBN film was detached from the Cu (111)/sapphire substrate by electrochemical delamination. A poly(methyl methacrylate) (PMMA) film was first spin-coated on the as-grown hBN/Cu (111)/sapphire as a protection layer. Then a thermal release tape (TRT; catalogue number 3195M from Nitto) was applied to the PMMA/hBN/Cu (111)/sapphire in order to avoid possible folding during the transfer process. Electrochemical delamination was performed using an aqueous solution of NaOH (1 M) as the electrolyte, with the Cu layer in the TRT/PMMA/hBN/Cu (111)/sapphire stack as the cathode and a platinum foil as the anode, under an applied DC voltage of 4 V. During this process, the TRT/PMMA/hBN stacked film was detached from the Cu (111)/sapphire through the generation of hydrogen bubbles at the hBN/Cu interface. After detachment, the TRT/PMMA/hBN stacked film can be placed on the target substrate. The TRT can be released by baking the TRT/PMMA/hBN/substrate on a hot-plate at 180 °C. The PMMA film was finally removed by immersing the sample in hot acetone for 40 min, leaving behind a monolayer hBN film on the target substrate.

Chemical vapour deposition of MoS_2

Highly oriented monolayer MoS_2 was grown on sapphire substrates by CVD in a horizontal hot-wall three-inch furnace tube with two heating zones. The maximum substrate size can be a wafer of up to two inches. High-purity sulfur (99.5%, Alfa) and MoO_3 (99%, Aldrich) powders were used as the reaction precursors. The sulfur powder was placed in the front heating zone at the upstream side of the furnace, and the temperature was maintained at 140 °C during the reaction. The MoO_3 powder was put into a quartz boat in the central heating zone of the furnace. The temperature of this central heating zone was gradually ramped to 740 °C and held for 5 min. During this process, MoS_2 was grown on the sapphire substrates placed at the downstream side of the MoO_3 quartz boat. All growth was performed in argon flowing gas (90 sccm) at a base pressure of 30 torr. Finally, the furnace was naturally cooled to room temperature¹⁷.

Transfer of monolayer MoS_2

After CVD growth, monolayer MoS_2 films on sapphire were transferred onto a target substrate using PMMA. PMMA was first

spin-coated onto the as-grown MoS_2 /sapphire as a protection layer. Then a TRT was applied to the PMMA/ MoS_2 /sapphire in order to avoid possible folding during the transfer process. The TRT/PMMA/ MoS_2 /sapphire stacked film was immersed in an NH_4OH solution (NH_4OH /deionized water, 17/100) at 100 °C for 20 min. Next, the detached TRT/PMMA/ MoS_2 stacked films were immersed in deionized water in order to dilute the etchant and residues. The TRT/PMMA/ MoS_2 stacked film is placed on the target substrate; the TRT is released by baking on a hot-plate at 180 °C; and the PMMA film is finally removed by immersing the sample in hot acetone for 40 min, leaving behind a monolayer MoS_2 film on the target substrate¹⁷.

LEEM/PEEM/ μ -LEED characterizations

We carried out structural analysis of hBN using low-energy electron microscopy (LEEM) with a field-emission gun, photoemission electron microscopy (PEEM) excited by a mercury lamp, and microspot low-energy electron diffraction (μ -LEED, with an area of about 3 μm in diameter). The system (Elmitec LEEM-III) is composed of a preparation chamber, a main chamber for imaging, and a deep ultraviolet laser¹⁸. The sample was first annealed in the preparation chamber in ultrahigh vacuum (1×10^{-9} torr) at 600 °C for 6 h and then transferred to the main chamber for LEEM/PEEM/ μ -LEED analysis.

STM characterizations

We used an Omicron ultrahigh vacuum (UHV) variable-temperature (VT)-STM system for atomic-scale structural characterization, with a base pressure better than 10^{-10} mbar. The hBN/Cu sample was annealed at 900 K for 5.5 h before STM measurements. All of the the STM images were captured under a constant-current mode at room temperature. The atomic-scale morphology of hBN/Cu was analysed directly by high-resolution STM.

EBSD characterization

EBSD was performed on a JEOL JSM-7800F PrimeSEM with collection accessory (from EDAX) at an accelerating voltage of 20 kV and the sample stage tilting at 70°.

TEM characterization

We obtained TEM images from transferred hBN/ SiO_2 /Si with an acceleration voltage of 200 keV using a FEI Tecnai Osiris transmission electron microscope.

AFM characterization

AFM data were acquired in tapping mode over the scan area, using a silicon tip.

XPS measurements

XPS spectra were obtained using a Perkin Elmer PHI 5400 system equipped with hemispherical analyser with an overall resolution of 0.05 eV. The energy span and energy scale were calibrated by setting the $4f_{7/2}$ line of gold and the $2p_{3/2}$ line of copper to 84 eV and 932.67 eV, respectively. XPS spectra were measured from the as-grown monolayer hBN film on Cu (111)/sapphire. The B 1s and N 1s emission peaks confirm the formation of hBN. The binding energies of B 1s and N 1s are located at 190.4 eV and 398 eV, respectively. The B/N atomic ratio is calculated from the integrated intensities of these peaks, yielding a ratio of 1/1.03, indicative of a good stoichiometry for the CVD-grown monolayer hBN films.

Raman measurements

The Raman measurements were carried out on transferred hBN film using a 532-nm solid-state laser as the excitation source. Excitation light with a power of 2.5 mW was focused onto the sample with

a $\times 100$ objective lens (with a numerical aperture of 0.9). The signal was collected using the same objective lens, analysed with a 0.75-m monochromator and detected with a liquid-nitrogen-cooled charge-coupled-device (CCD) camera. The E_{2g} band of hBN is located at $1,371\text{ cm}^{-1}$ and its FWHM value is 14.5 cm^{-1} , confirming that the transferred hBN film is monolayer thick with high crystalline quality.

X-ray diffraction measurements

X-ray diffraction (XRD; Bruker D8-Discover) θ – 2θ and 360° azimuthal (φ) scans were conducted using a Cu K α radiation source ($\lambda = 1.54\text{ \AA}$). An XRD azimuthal φ scan of twinned Cu (111) films was operated at $\chi = 70.5^\circ$, $\omega = 16.2^\circ$ and $2\theta = 43.3^\circ$; an XRD azimuthal φ scan of untwinned Cu (111) films was operated at $\chi = 70.5^\circ$, $\omega = 21.5^\circ$ and $2\theta = 43.3^\circ$.

First-principles calculations

First-principles calculations were carried out using DFT as implemented in the Vienna ab initio simulation package (VASP)¹⁹ within MedeA software²⁰. We used the projector augmented wave method, the exchange–correlation potential described by the Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation (GGA)²¹, and van der Waals correction vdW-DF (optB86b) functionals²² in order to calculate the distance and binding energy between BN molecules and copper substrates. For the plane-to-plane and edge-to-step epitaxy model, a rigid B_6N_7 molecule is physically absorbed onto a 6×6 and a $6 \times 7\sqrt{3}$ Cu (111) surface with a thickness of four copper layers, using a $2 \times 2 \times 1$ and a $2 \times 1 \times 1$ k -grid, respectively. To eliminate spurious interactions resulting from the slab model of the supercell, the vacuum thickness is larger than 17 \AA , the in-plane distance of a B_6N_7 molecule away from another side of the step edge is at least 10 \AA , and the energy cutoff of the plane waves is 400 eV . The lattice constant of Cu (111) used in simulation is $a = 2.6\text{ \AA}$ at a reaction temperature of $1,000^\circ\text{C}$, with hBN assuming the same lattice constant for simplification. The interlayer distance of Cu (111) is 2.05 \AA . The height of the BN molecule from the Cu surface is optimized until the change in the energy and the force reach 10^{-4} eV and $10^{-2}\text{ eV \AA}^{-1}$, respectively. For plane-to-plane epitaxy, the optimized vertical distances from the top layer of the Cu surface, d , for six configurations are $d = 1.98\text{ \AA}$ for N_1B_{III} , $d = 1.77\text{ \AA}$ for $N_{II}B_{III}$, $d = 2.03\text{ \AA}$ for $N_{III}B_I$, $d = 1.98\text{ \AA}$ for N_1B_{III} , $d = 1.77\text{ \AA}$ for $N_{III}B_{III}$, and $d = 2.03\text{ \AA}$ for $N_{II}B_I$. Binding energies are defined by $-E_b = E_{B_6N_7-Cu} - E_{B_6N_7} - E_{Cu}$, where $E_{B_6N_7-Cu}$ is the total energy of the hBN flake/Cu (111) system, and $E_{B_6N_7}$ and E_{Cu} are the energies of the B_6N_7 flake and Cu substrate (with or without a step edge); the calculated binding energies of six total stacking structures in the presence of a step are plotted by varying the in-plane distance, D_i , to the step edge. The six lowest energetic configurations in the presence of a step are located at $D_i = 2a/\sqrt{3}$ for N_1B_{III} ($d = 1.99\text{ \AA}$), $4a/\sqrt{3}$ for $N_{II}B_{III}$ ($d = 1.77\text{ \AA}$), $1.5a/\sqrt{3}$ for $N_{III}B_I$ ($d = 2.08\text{ \AA}$), $2.5a/\sqrt{3}$ for N_1B_{III} ($d = 1.98\text{ \AA}$), $1.5a/\sqrt{3}$ for $N_{III}B_{III}$ ($d = 1.84\text{ \AA}$), and $2a/\sqrt{3}$ for $N_{II}B_I$ ($d = 2.06\text{ \AA}$) (Extended Data Fig. 8). Note that when D_i is less than $\sqrt{3}a/3$ (too close to the step-edge), the energy dramatically increases.

We study here the energy difference between N_1B_{III} (0°) and N_1B_{II} (60°), the two lowest-energy structures docking to A-step and B-step edges, with different lengths of docking contact. We construct a model of a BN stripe composed of aromatic rings docking to the step edge. We find that the energy difference between the two configurations N_1B_{III} (0°) and N_1B_{II} (60°)—an indicator of selectivity for mono-orientation—increases rapidly with the docking contact, and approaches a δE of roughly 0.78 eV , amplifying the Boltzmann selectivity factor $e^{\delta E/k_B T}$ (with a $k_B T$ of around 0.11 eV at $1,300\text{ K}$) for a contact length of just five to six hexagons to more than 10^3 (Extended Data Fig. 9a, b). The large energy difference means that the step edge plays the role of ‘orientation filter’, allowing a predominant phase of BN (a mono-orientation) during growth. Factually, the competing subcritical nuclei must be larger and, especially, are probably elongated along the step-edge to allow a more energetically favourable contact. Obviously such an energy value

roughly scales with the contact length, C , and the selectivity factor increases very rapidly with size, as $\exp(C)$.

The effect of misfit and misalignment of BN to the step edge can be studied by calculating the binding energy of a B_7N_7 molecule at a three-hexagon contact length with a small tilt angle along the step edge (the aligned structure with a tilt angle of zero corresponds to N_1B_{III} and N_1B_{II}). The lattice constants of Cu (111) and hBN are allowed to be different as $a_{Cu(111)} = 2.6\text{ \AA}$ and $a_{hBN} = 2.5\text{ \AA}$, with a 3.8% lattice misfit. We find that, although the total binding energy is weakened because of lattice misfit (that is, there is a weaker interaction of plane-to-plane epitaxy), the energy difference between the 0° and 60° orientations remains nearly the same (from 0.3 eV to 0.27 eV). The calculated binding weakens as the tilted angle becomes larger, indicating that the most stable configuration during the initial nucleation growth occurs when docking is tight and well aligned to the step edge (Extended Data Fig. 9c, d). The result suggests that mono-orientated hBN was epitaxially grown on Cu (111) in a manner that was guided mostly by step edges, and that the moiré was formed to release the strain between Cu (111) and hBN (lattice mismatch or surface topography) after a large area of film had grown. We note that the cooling step after large-area hBN growth will result in strains that might be released by local straining of hBN and formation of a moiré pattern. Therefore, we believe that moiré does not actually affect the growth process, but occurs afterwards.

Data availability

All data needed to evaluate our conclusions are found in the main text and the Extended Data. Further data related to the paper are available from the corresponding authors on reasonable request.

- Hsu, W.-F. et al. Monolayer MoS_2 enabled single-crystalline growth of AlN on Si(100) using low-temperature helicon sputtering. *ACS Appl. Nano Mater.* **2**, 1964–1969 (2019).
- Jin, L., Fu, Q., Mu, R., Tan, D. & Bao, X. Pb intercalation underneath a graphene layer on Ru(0001) and its effect on graphene oxidation. *Phys. Chem. Chem. Phys.* **13**, 16655–16660 (2011).
- Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
- MedeA (Materials Design Inc, 2016).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- Klimeš, J., Bowler, D. R. & Michaelides, A. Van der Waals density functionals applied to solids. *Phys. Rev. B* **83**, 195131 (2011).
- Yankowitz, M. et al. Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nat. Phys.* **8**, 382–386 (2012).
- Joshi, S. et al. Boron nitride on Cu (111): an electronically corrugated monolayer. *Nano Lett.* **12**, 5821–5828 (2012).

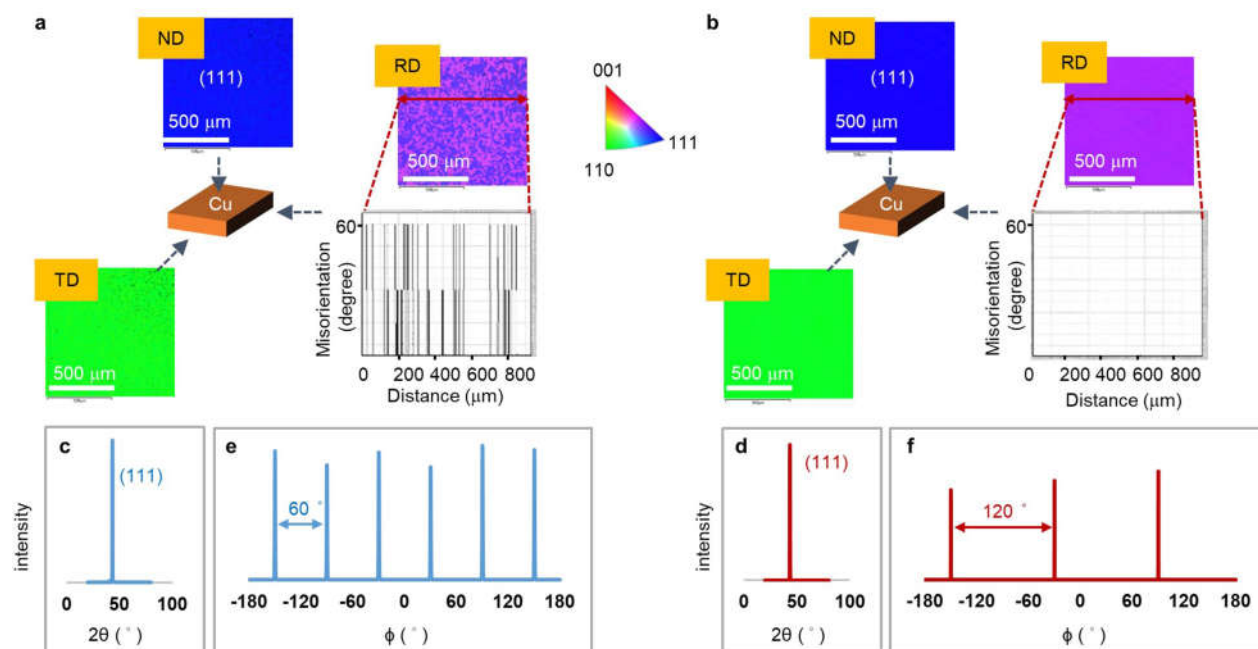
Acknowledgements Tse-An Chen, C.-P.C., H.-S.P.W. and L.-J.L. acknowledge support from the Taiwan Semiconductor Manufacturing Company (TSMC). W.-H.C. acknowledges support from the Ministry of Science and Technology of Taiwan (grants MOST-108-2119-M-009-011-MY3 and MOST-107-2112-M-009-024-MY3) and from the CEFMS of the National Chiao Tung University, supported by the Ministry of Education of Taiwan. Y.Z. acknowledges financial support from the National Natural Science Foundation of China (grant 51861135201). Q.F. thanks the National Natural Science Foundation of China (grants 21688102 and 21825203) and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB17020000) for financial support. B.I.Y. acknowledges support from the US Department of Energy (grant DE-SC0012547) and a stimulating discussion with T. Ivanov (US Army Research Laboratory). Tse-An Chen and L.-J.L. acknowledge useful discussions with S. Brems at Imec.

Author contributions L.-J.L. and Tse-An Chen conceived the project. Tse-An Chen, C.-C.T. and C.-K.W. grew the hBN by CVD, performed the transfer of hBN, and carried out EBSD, Raman and AFM measurements. C.-P.C. performed first-principles calculations, and C.-P.C. and B.I.Y. carried out theoretical analysis. R.L. and Q.F. performed μ -LEED measurements. Y.Z. and S.P. performed STM experiments. C.-K.W., Tzu-Ang Chao and W.-C.C. fabricated the metal–insulator–metal device and the MoS_2 FET. Tse-An Chen, C.-K.W. and Tzu-Ang Chao performed electrical measurements. L.-J.L., W.-H.C. and H.-S.P.W. supervised the project. All of the authors discussed the results and wrote the paper.

Competing interests The authors declare no competing interests.

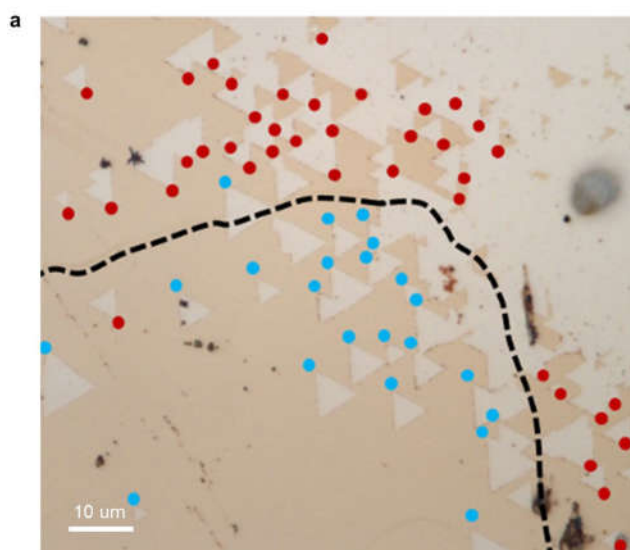
Additional information

Correspondence and requests for materials should be addressed to B.I.Y., W.-H.C. or L.-J.L.
Reprints and permissions information is available at <http://www.nature.com/reprints>.

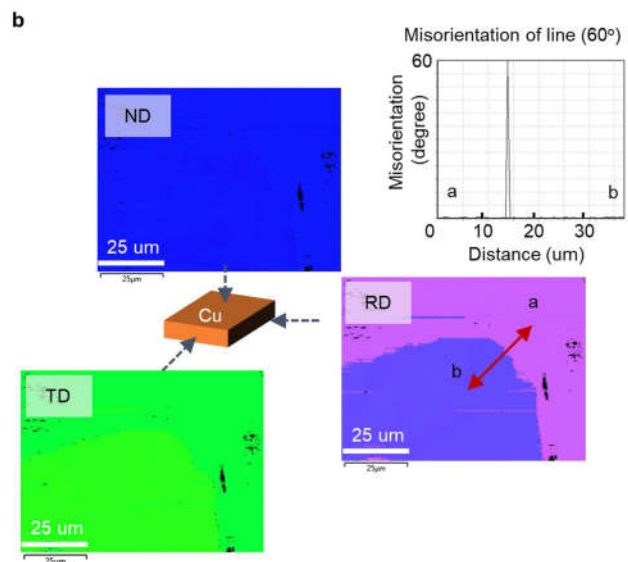


Extended Data Fig. 1 | Analysis of Cu (111) crystal orientation on c-sapphire substrates. **a**, EBSD inverse pole figure (IPF) mapping (1 mm \times 1 mm) of the Cu substrate annealed at 1,000 °C for 1 h. The normal direction (ND), transverse direction (TD) and rolling direction (RD) mappings, as indicated by the triangular colour map, show that Cu (111) is polycrystal. The line scan of misorientation on the RD map indicates an in-plane 60° rotation. **b**, IPF map of the Cu substrate annealed at 1,050 °C for 1 h; the film is characterized as single-crystal Cu (111), and no twinned grain is founded. **c**, **d**, XRD θ - 2θ scans of the

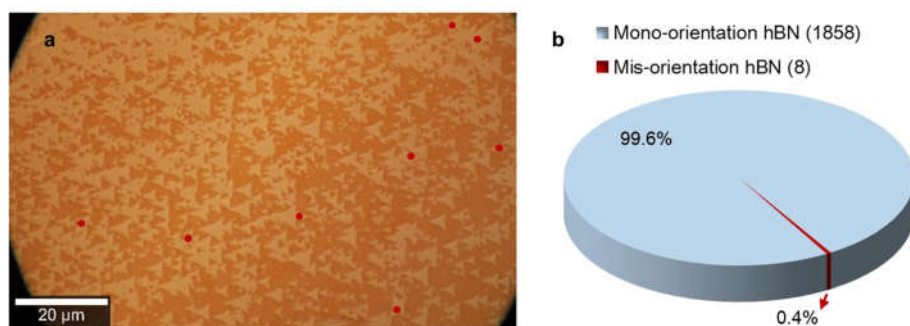
Cu (111)/c-sapphire substrate annealed at 1,000 °C for 1 h (**c**) and annealed at 1,050 °C for 1 h (**d**), revealing a Cu (111) peak at $2\theta = 43.3^\circ$. **e**, **f**, XRD ϕ scans of the Cu (111)/c-sapphire substrate annealed at 1,000 °C for 1 h with an in-plane rotation of 60° (**e**) and annealed at 1,050 °C for 1 h with an in-plane rotation of 120° (**f**). Note that the sample annealed at 1,050 °C shows the signature of a single crystal without in-plane misorientation, because an hBN triangle has C_3 symmetry and is symmetric after a 120° rotation.



Extended Data Fig. 2 | Single-oriented hBN flakes in different Cu grains.
a, Optical micrograph of hBN molecules (triangles) grown on different Cu (111) grains (the black dashed line indicates the boundary between adjacent twin grains). Oppositely oriented hBN flakes are marked by red and blue circles.

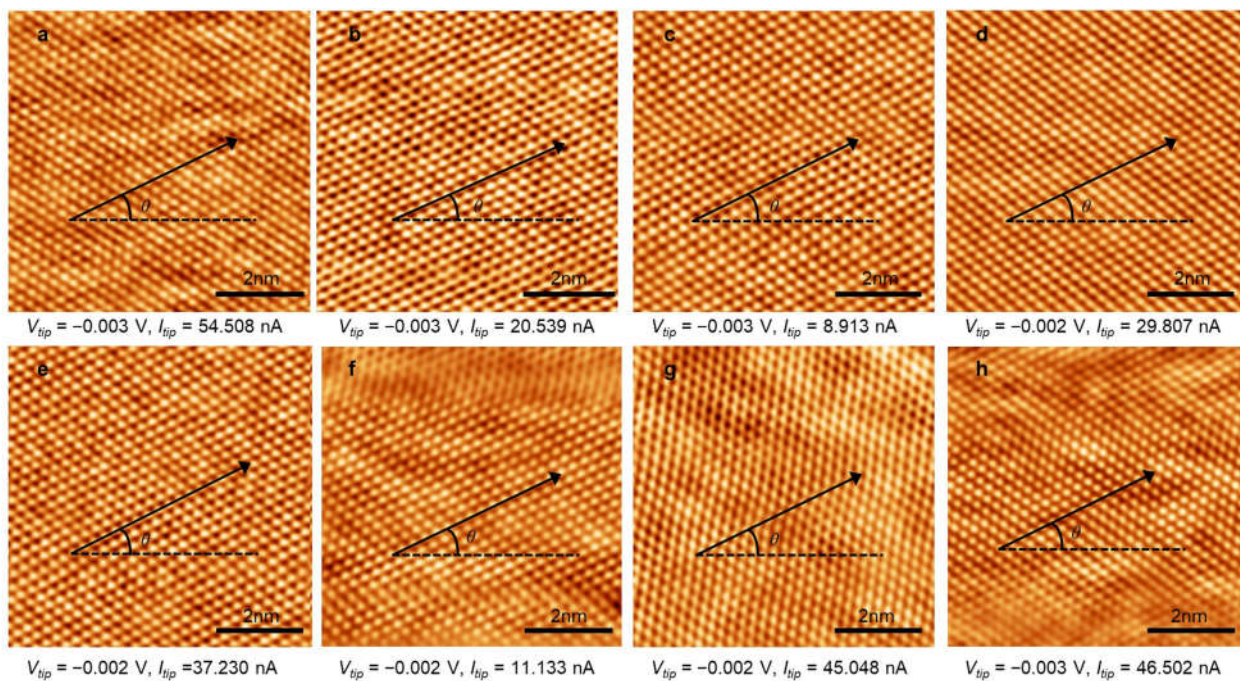


b, EBSD IPF maps of the area shown in **a**. The misorientation-of-line scan indicates that the twin grain is rotated by 60° (grain a to grain b). The RD map clearly shows the difference in in-plane orientation between grains a and b.

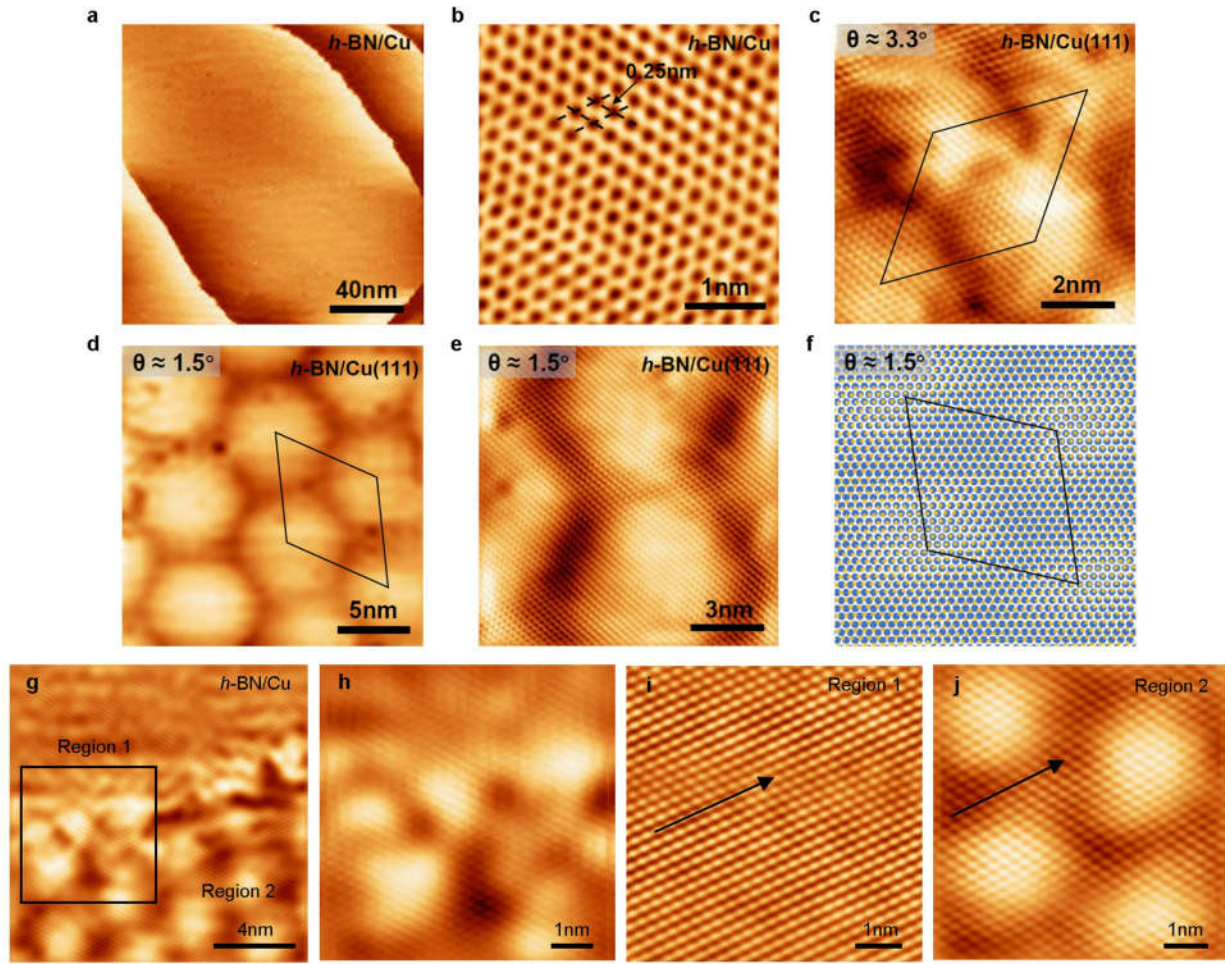


Extended Data Fig. 3 | Statistical analysis of the orientation distribution of triangular hBN flakes. **a**, Optical micrograph of hBN grown on a Cu (111)/c-sapphire substrate at 1,050 °C. Misaligned hBN flakes are marked by red

circles. **b**, Statistical analysis of the optical micrograph from **a**; more than 99.6% of the hBN flakes are aligned in one direction on Cu (111).

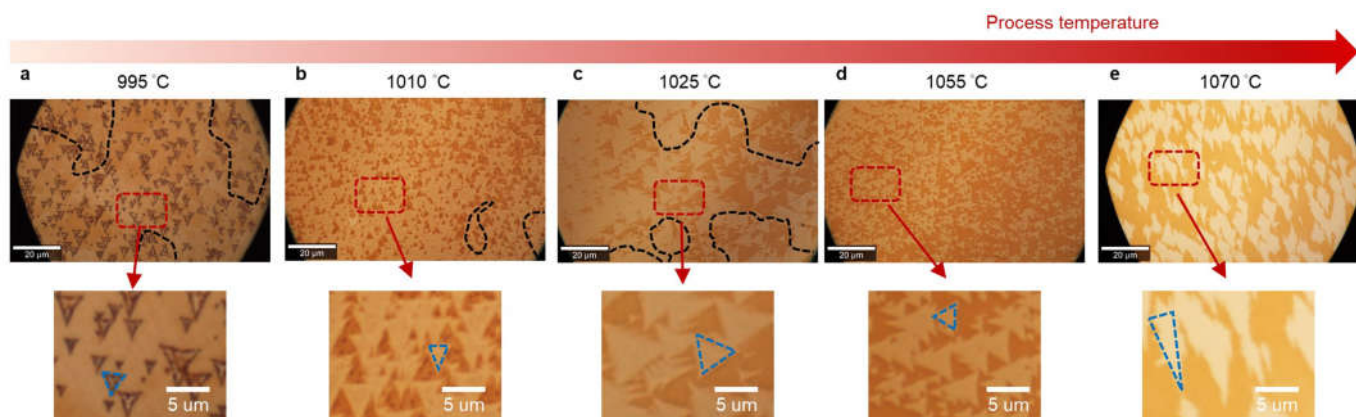


Extended Data Fig. 4 | STM images from randomly selected locations. a–h, Images of a $1.5 \times 1.5 \text{ cm}^2$ hBN film grown on Cu(111)/c-sapphire, where the angle (θ) between the hBN lattice orientation (black arrow) and the horizontal line (black dashed line) in each image is $26.5^\circ \pm 1^\circ$, indicating that the hBN is a single-crystal film.



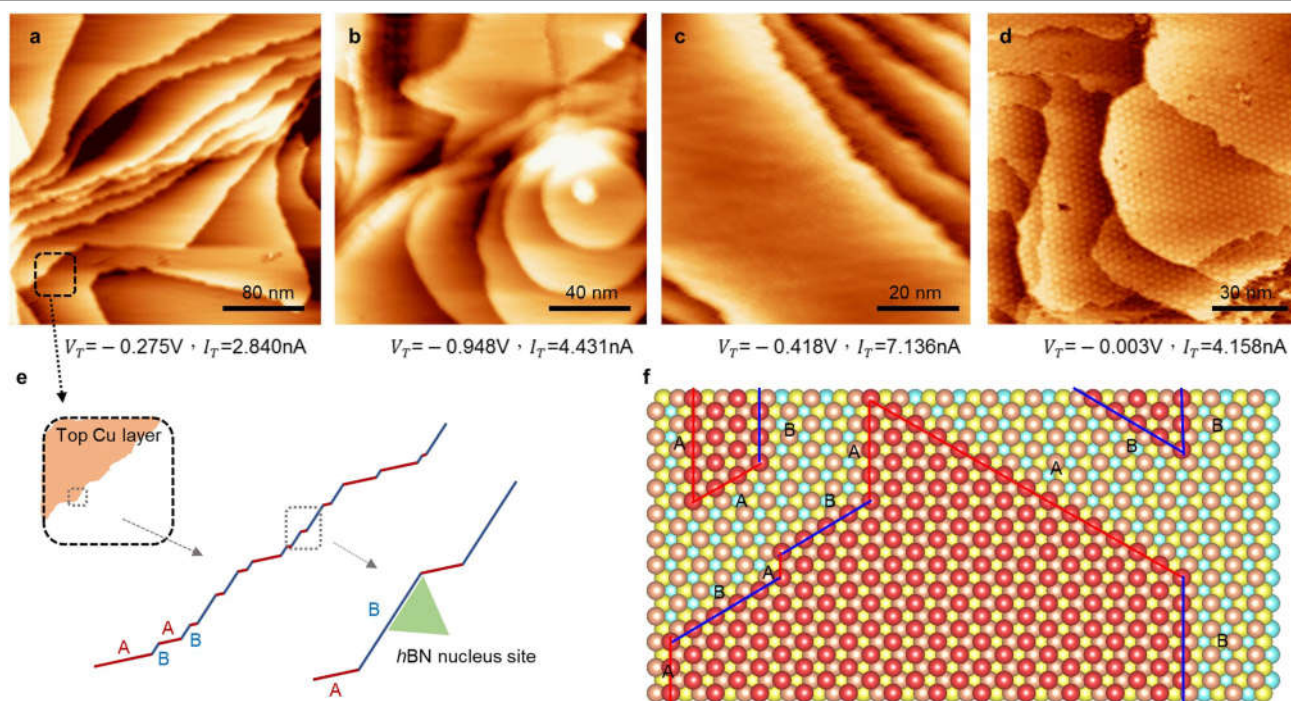
Extended Data Fig. 5 | STM characterization of the atomic structure of monolayer hBN on Cu (111). **a**, Large-scale STM image ($V_{\text{tip}} = -1.008$ V; $I_{\text{tip}} = 3.90$ nA; $T = 300$ K) of hBN/Cu. **b**, Atomic-scale STM image ($V_{\text{tip}} = -0.003$ V; $I_{\text{tip}} = 46.50$ nA; $T = 300$ K) of hBN/Cu. **c**, Typical STM image ($V_{\text{tip}} = -0.003$ V; $I_{\text{tip}} = 8.91$ nA; $T = 300$ K) of hBN on Cu (111) with a relative rotation angle, θ , of approximately 3.3° , showing a moiré pattern with a period of 4.20 nm. **d**, Typical STM image ($V_{\text{tip}} = -0.032$ V; $I_{\text{tip}} = 18.51$ nA; $T = 300$ K) of hBN on Cu (111) with a θ of roughly 1.5° , showing a moiré pattern with a period of 7.75 nm. The unit cell of the moiré pattern is highlighted by a black rhombus. **e**, Magnified STM image ($V_{\text{tip}} = -0.039$ V; $I_{\text{tip}} = 18.51$ nA; $T = 300$ K) of hBN on Cu (111) with a θ of approximately 1.5° . **f**, Simulation of the moiré pattern for monolayer hBN on Cu (111) with a θ of roughly 1.5° . The unit cell of the moiré pattern for hBN/Cu (111) is highlighted by a black rhombus. The large-scale STM image in **a** shows a large-area flat terrace of hBN/Cu with a clean surface. The atomic-scale STM image in **b** reveals a honeycomb structure with a lattice constant of roughly 0.25 nm, which coincides well with the lattice parameters of hBN. Notably, in some typical regions of hBN/Cu (111), moiré patterns with different periods are observed. For instance, **d** and **e** show a moiré pattern with a period of around 7.75 nm, and **c** shows another with a period of roughly 4.20 nm. Such patterns arise from the lattice mismatch and/or relative rotation between hBN and the underlying Cu (111) substrate, and the moiré periods (D) correlate with the relative rotation angles (θ) between hBN and Cu (111) as²³ $D = (1 + \delta)a / \sqrt{2(1 + \delta)(1 - \cos\theta)} + \delta^2$, where δ is the lattice mismatch (roughly 2%) between hBN and the Cu (111) lattice²⁴, and a is the lattice constant of hBN.

Consequently, the θ for the moiré pattern with a D of around 7.75 nm (**d, e**) is calculated to be around 1.5° , and the simulated moiré pattern generated from monolayer hBN stacking on Cu (111) with a θ of roughly 1.5° (**f**) fits well with the STM result (**d, e**). The θ for the moiré pattern with a D of around 4.20 nm is calculated to be roughly 3.3° (**c**). **g–j**, STM images showing the boundary between areas with and without moiré pattern. **g**, Typical STM image ($V_{\text{tip}} = -0.003$ V; $I_{\text{tip}} = 8.10$ nA; $T = 300$ K) of hBN/Cu at the boundary site. **h**, Magnified STM image ($V_{\text{tip}} = -0.003$ V; $I_{\text{tip}} = 8.10$ nA; $T = 300$ K) of the boundary in **g** (highlighted by the black square), showing that the hBN lattices present perfect coherence at the boundary site. **i**, Magnified STM image ($V_{\text{tip}} = -0.003$ V; $I_{\text{tip}} = 17.93$ nA; $T = 300$ K) of hBN/Cu without moiré pattern in region 1 of **g**. **j**, Magnified STM image ($V_{\text{tip}} = -0.003$ V; $I_{\text{tip}} = 10.78$ nA; $T = 300$ K) of hBN/Cu with moiré pattern in region 2 of **g**. The hBN atomic rows in region 1 and region 2 are along the same direction (black arrows). The STM image in **g** was captured at a typical boundary region, with moiré (region 2) and non-moiré (region 1) areas. The magnified atomic-resolution image at the boundary site in **h** shows that the hBN presents perfect lattice coherence at the patching boundary. The hBN atomic rows in the two adjacent regions are along the same direction (**i, j**). All images suggest that the hBN is aligned well and has mono-orientation, indicating epitaxial growth of hBN on Cu (111), and that the formation of moiré pattern does not affect the hBN orientation. We believe that the hBN completes its single-crystal growth at high temperatures, and that the strain associated with sample cooling after growth results in local moiré pattern.



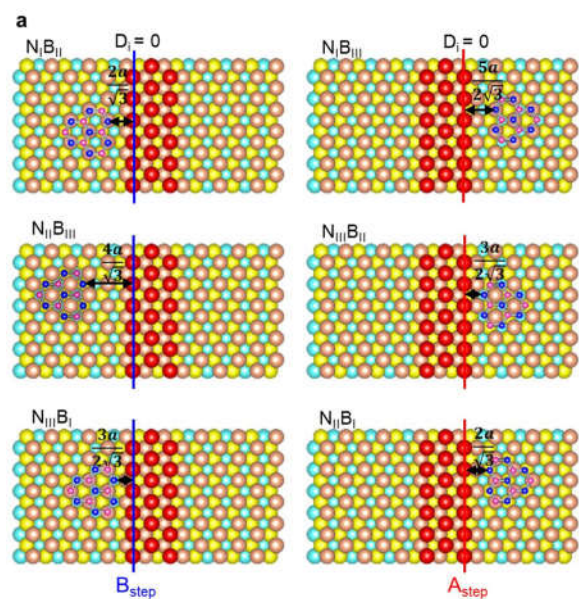
Extended Data Fig. 6 | Optical micrograph of hBN grown on Cu (111)/ c-sapphire at different temperatures. All as-grown hBN/Cu (111) was slowly oxidized at 150 °C in air in order to enhance the contrast between hBN and Cu. Black dashed lines indicate Cu twin grain boundaries. Regions with a red dashed outline are magnified in lower images. **a**, hBN grown at 995 °C, with hBN triangles aligned in the same Cu grain. **b**, hBN grown at 1,010 °C. Note that the

hBN flakes grown at 995 °C and 1,010 °C are easily damaged after oxidation at 150 °C in air. **c**, **d**, All hBN triangles point in the same direction in the same Cu (111) grain after growth at 1,025 °C and 1,055 °C, and triangles seem unchanged after oxidation. **e**, After growth at high temperature (1,070 °C), all hBN flakes align in the same direction, but show different (stretched) shapes.



Extended Data Fig. 7 | STM images of hBN/Cu(111). a–d, STM images of hBN/Cu steps without moiré pattern (a–c) and with moiré pattern (d), showing that step edges are often observed in our Cu(111) crystals. e, Diagram showing that the meandering steps consist of segments of A and B types, and that BN

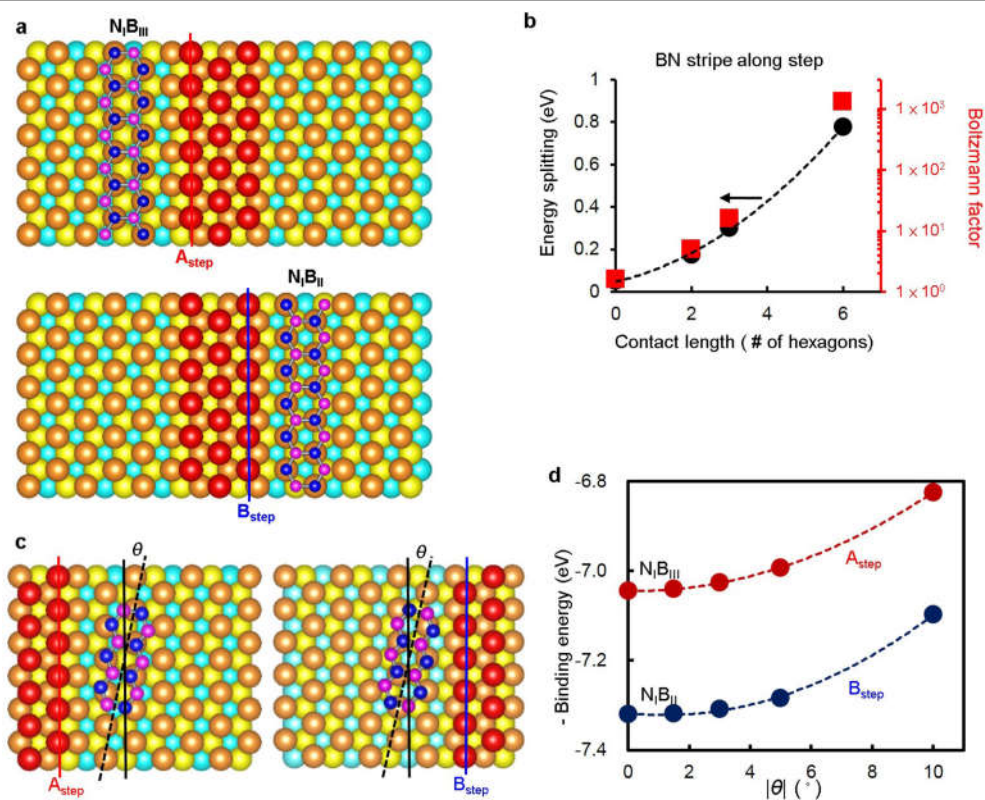
kinetically nucleates at B-to-A corners while docking to stronger binding sites (B steps) with proper orientation. f, Atomic model of step edges on the top Cu(111) surface, showing A and B steps.



Extended Data Fig. 8 | Calculated binding energy of six typical B_6N_7 -Cu (111) configurations, taking into account docking to top-layer Cu step edges.

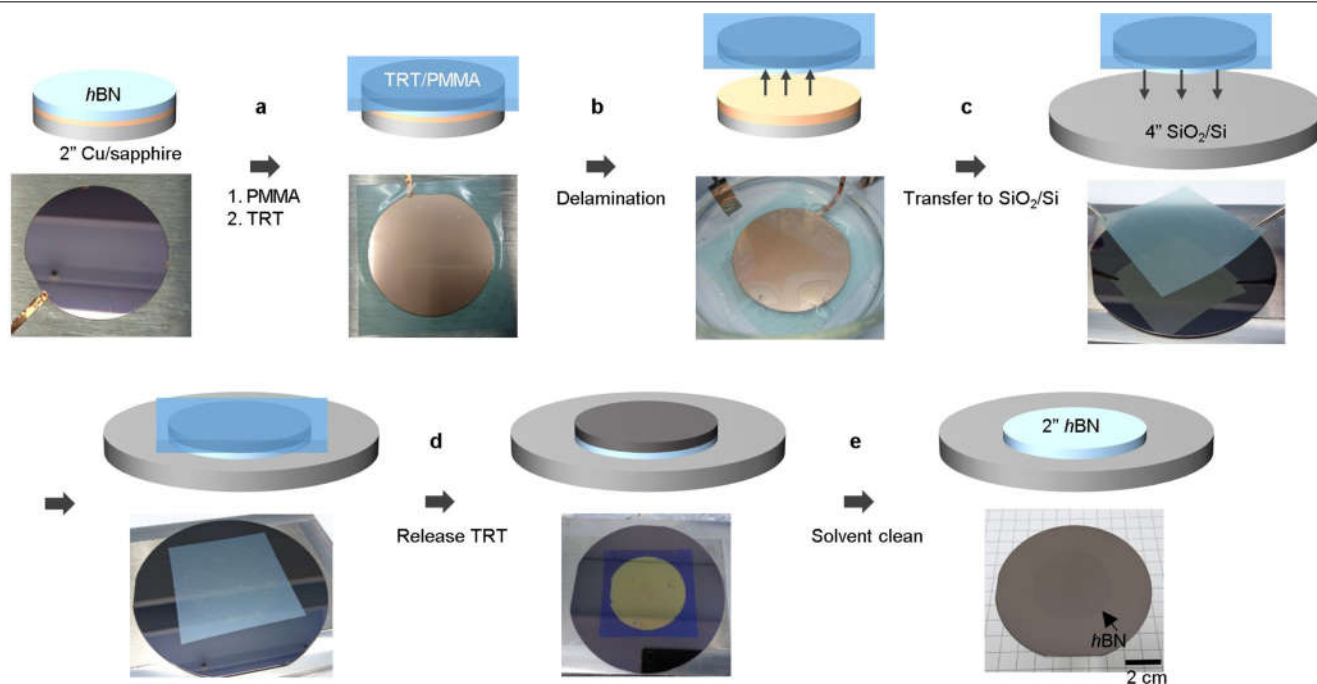
a, Atomic models of six configurations, showing the optimized in-plane distance, D_i , away from A-step (red) or B-step (blue) edges. **b**, Binding energies

as a function of D_i (in units of $\sqrt{3}a/3$, where a is the lattice constant of Cu (111)), with the lowest-energy configurations marked with arrows. Plane-to-plane epitaxy corresponds to the absence of (or infinity distance away from) the step



Extended Data Fig. 9 | Comparison of planar epitaxy and step-edge docking of hBN on Cu (111). **a**, Atomic models of fully docking configurations, N_7B_{II} (0°) and N_7B_{III} (60°). **b**, Calculated binding-energy differences (left axis, black circles) and corresponding Boltzmann factors (right axis, red squares) between N_7B_{II} and N_7B_{III} for stripes of BN on Cu (111) as a function of docking length (in number of hexagons) on the top-layer Cu step edge. **c**, **d**, Calculated energy of misfit B_7N_7 -Cu (111) configurations as a function of small tilted angle along with

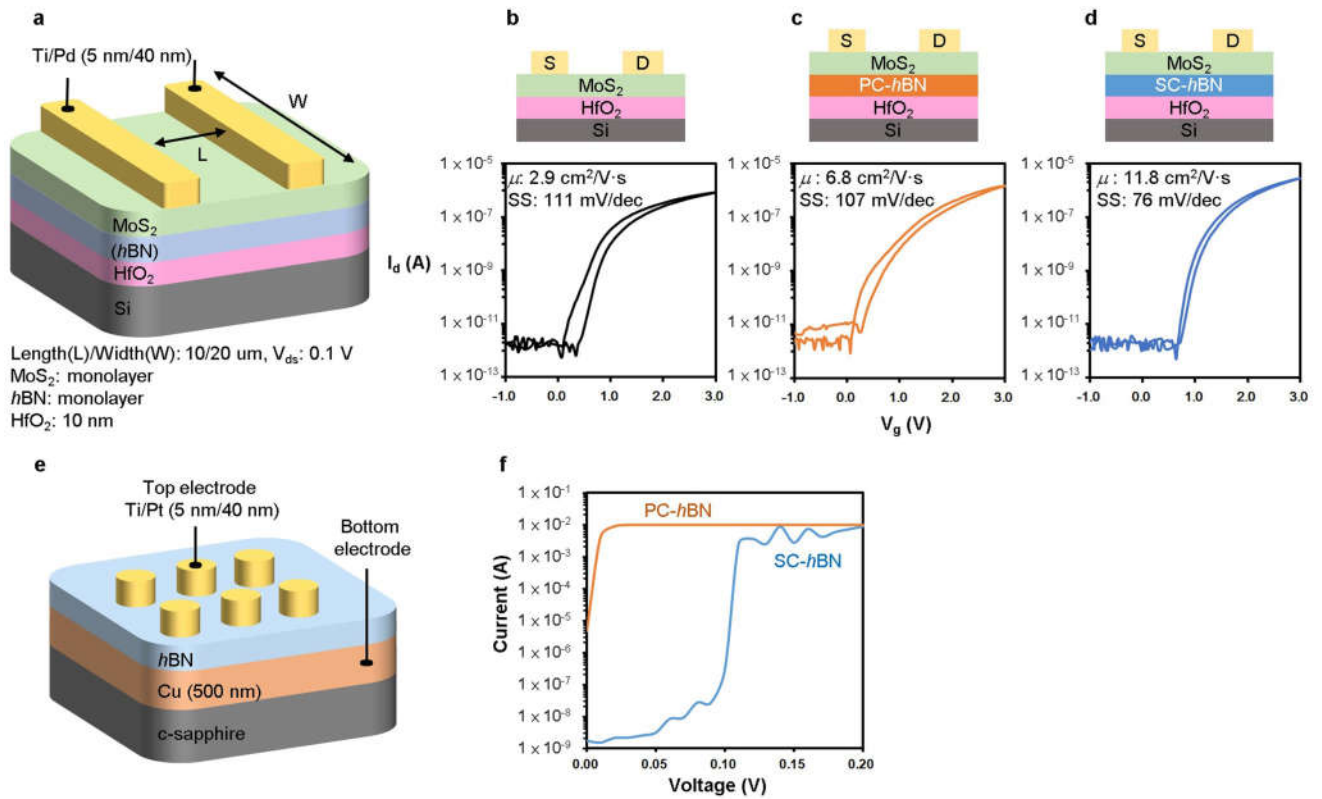
top-layer Cu step-edge docking. **c**, Atomic models of B_7N_7 -Cu (111) configurations docking to A steps or B steps at a tilt angle of θ . The aligned structures ($\theta = 0^\circ$) correspond to N_7B_{II} and N_7B_{III} (the two lowest-energy structures docking in the vicinity of A-step and B-step edges). The lattice misfit between hBN and Cu (111) is 3.8% ($a_{hBN} = 2.5 \text{ \AA}$; $a_{Cu(111)} = 2.6 \text{ \AA}$). **d**, Binding varies with the tilted angle along the step edge.



Extended Data Fig. 10 | Diagrams and photographs illustrating hBN transfer.

a, A PMMA film is first spin-coated on the as-grown hBN/Cu (111)/sapphire as a protection layer. A TRT is then applied to the PMMA/hBN/Cu (111)/sapphire in order to avoid folding during the transfer process. **b**, Electrochemical delamination is carried out using an aqueous solution of NaOH (1M) as the electrolyte, the Cu layer in the TRT/PMMA/hBN/Cu (111)/sapphire stack as the cathode, and a platinum foil as the anode, with an applied DC voltage of 4 V.

During this process, the TRT/PMMA/hBN stacked film is detached from the Cu (111)/sapphire through the generation of hydrogen bubbles at the hBN/Cu interface. **c**, The TRT/PMMA/hBN stacked film is placed on the four-inch SiO₂/Si substrate. **d**, TRT can be released by baking the TRT/PMMA/hBN/substrate on a hot-plate at 180 °C. **e**, The PMMA film is removed by immersing the sample in hot acetone for 40 min, leaving behind a two-inch monolayer hBN film on the four-inch SiO₂/Si wafer.



Extended Data Fig. 11 | FETs built with and without hBN as an interface dielectric in a bottom-gate configuration. **a–d**, Diagrams showing the various MoS₂ FET device structures examined and their transfer characteristics (drain current, I_d , versus gate voltage, V_g) at a driving voltage (V_{ds}) of 0.1 V. Monolayer hBN and MoS₂ are synthesized by CVD (see Methods). **a**, The hBN film is transferred onto a Si substrate with a 10-nm HfO₂ dielectric layer, and topped with a single layer of MoS₂. The metal contacts are processed by photolithography, and evaporated with Ti (5 nm) and Pd (50 nm) using an e-gun evaporator. **b–d**, Top, diagrams showing a typical monolayer MoS₂ FET on HfO₂ (**b**), an MoS₂ monolayer on a polycrystalline (PC) hBN monolayer (**c**), and an MoS₂ monolayer on a single-crystal (SC) hBN monolayer (**d**). Bottom, I_d versus V_g . The extracted two-probe electron mobility at room temperature in vacuum (μ) is 2.9 $\text{cm}^2/\text{V}\cdot\text{s}$, 6.8 $\text{cm}^2/\text{V}\cdot\text{s}$ and 11.8 $\text{cm}^2/\text{V}\cdot\text{s}$, respectively. The mobility of MoS₂ improves substantially (by about an order of magnitude) by replacing

polycrystalline hBN with a single-crystal hBN. A suppressed current hysteresis is also observed when using single-crystal hBN as the buffer layer, and the subthreshold swing (SS) also improved from 111 $\text{mV}\cdot\text{dec}^{-1}$ to 76 $\text{mV}\cdot\text{dec}^{-1}$. The results indicate that single-crystal hBN can be used to enhance the electrical performance of 2D-based transistors. **e**, Diagram showing the hBN metal-insulator-metal (MIM) structure used to examine the quality of hBN. **f**, I - V curves for MIM tunnel junctions with either a polycrystalline or a single-crystal hBN monolayer sandwiched between Pt/Ti (top) and Cu (bottom) electrodes. Electrical contacts were fabricated by photolithography and e-gun evaporation of Ti (5 nm) and Pt (40 nm) to form $100 \times 100 \mu\text{m}^2$ pads on as-grown hBN/Cu/c-sapphire substrate. The device with single-crystal monolayer hBN exhibits a large breakdown voltage (of around 0.1 V), whereas the device with polycrystalline monolayer hBN shows direct tunnelling characteristics.

In situ NMR metrology reveals reaction mechanisms in redox flow batteries

<https://doi.org/10.1038/s41586-020-2081-7>

Received: 6 June 2019

Accepted: 11 December 2019

Published online: 2 March 2020

 Check for updates

Evan Wenbo Zhao¹, Tao Liu^{1,5}, Erlendur Jónsson^{1,2}, Jeongjae Lee^{1,6}, Israel Temprano¹, Rajesh B. Jethwa¹, Anqi Wang³, Holly Smith¹, Javier Carretero-González⁴, Qilei Song³ & Clare P. Grey¹✉

Large-scale energy storage is becoming increasingly critical to balancing renewable energy production and consumption¹. Organic redox flow batteries, made from inexpensive and sustainable redox-active materials, are promising storage technologies that are cheaper and less environmentally hazardous than vanadium-based batteries, but they have shorter lifetimes and lower energy density^{2,3}. Thus, fundamental insight at the molecular level is required to improve performance^{4,5}. Here we report two in situ nuclear magnetic resonance (NMR) methods of studying redox flow batteries, which are applied to two redox-active electrolytes: 2,6-dihydroxyanthraquinone (DHAQ) and 4,4'-((9,10-anthraquinone-2,6-diyl)dioxy)dibutyrates (DBEAQ). In the first method, we monitor the changes in the ¹H NMR shift of the liquid electrolyte as it flows out of the electrochemical cell. In the second method, we observe the changes that occur simultaneously in the positive and negative electrodes in the full electrochemical cell. Using the bulk magnetization changes (observed via the ¹H NMR shift of the water resonance) and the line broadening of the ¹H shifts of the quinone resonances as a function of the state of charge, we measure the potential differences of the two single-electron couples, identify and quantify the rate of electron transfer between the reduced and oxidized species, and determine the extent of electron delocalization of the unpaired spins over the radical anions. These NMR techniques enable electrolyte decomposition and battery self-discharge to be explored in real time, and show that DHAQ is decomposed electrochemically via a reaction that can be minimized by limiting the voltage used on charging. We foresee applications of these NMR methods in understanding a wide range of redox processes in flow and other electrochemical systems.

Improved characterization methods of studying redox flow batteries are needed to enhance the performance and extend the battery life of both organic- and inorganic-based chemistries. Ex situ characterization can be challenging, owing to the high reactivity, sensitivity to sample preparation and short lifetimes of some of the oxidized and/or reduced redox-active molecules and ions within the electrolytes. However, one of the distinct features of redox flow batteries is the decoupling of their energy storage and power generation, which provides different opportunities for in situ monitoring. So far, methods such as in situ optical spectrophotometry⁶ and electron paramagnetic resonance (EPR)⁷ have been used to study, for example, crossover of quinones and vanadyl ions, but considerable opportunities remain to improve characterization methods to address limitations inherent to each method and to probe different phenomena. NMR spectroscopy has previously been used to study benzoquinone and polyoxometalate redox reactions in an in situ static electrochemical cell^{8–10}. Here we proceed a step further by using NMR to study species in flow via two different methods: probing

the electrolyte in the flow path (on-line detection), or in the battery cell (operando detection).

The two in situ NMR setups

In on-line detection (Fig. 1a and Extended Data Fig. 1a), the battery is positioned outside the NMR magnet (300 MHz) and one electrolyte solution is pumped through a flow apparatus in the NMR probe, enabling the study of either the catholyte or the anolyte. The setup requires minimum modification of a laboratory-scale flow battery and can be easily adapted to other solution NMR instruments and coupled with other analytical (flow) characterization methods. For operando detection (Fig. 1b and Extended Data Fig. 1b), a miniaturized flow battery cell is positioned inside the detection region of the NMR probe, enabling the simultaneous study of the catholyte and anolyte in the battery cell. The majority of the data presented below are acquired with the on-line detection scheme unless otherwise noted, owing to its higher sensitivity—the on-line setup has a larger sampling volume (7.3 cm³) than the

¹Department of Chemistry, University of Cambridge, Cambridge, UK. ²Department of Physics, Chalmers University of Technology, Gothenburg, Sweden. ³Barrar Centre, Department of Chemical Engineering, Imperial College London, London, UK. ⁴Institute of Polymer Science and Technology, ICTP-CSIC, Madrid, Spain. ⁵Present address: Shanghai Key Laboratory of Chemical Assessment and Sustainability, Department of Chemistry, Tongji University, Shanghai, China. ⁶Present address: School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea. ✉e-mail: cpg27@cam.ac.uk

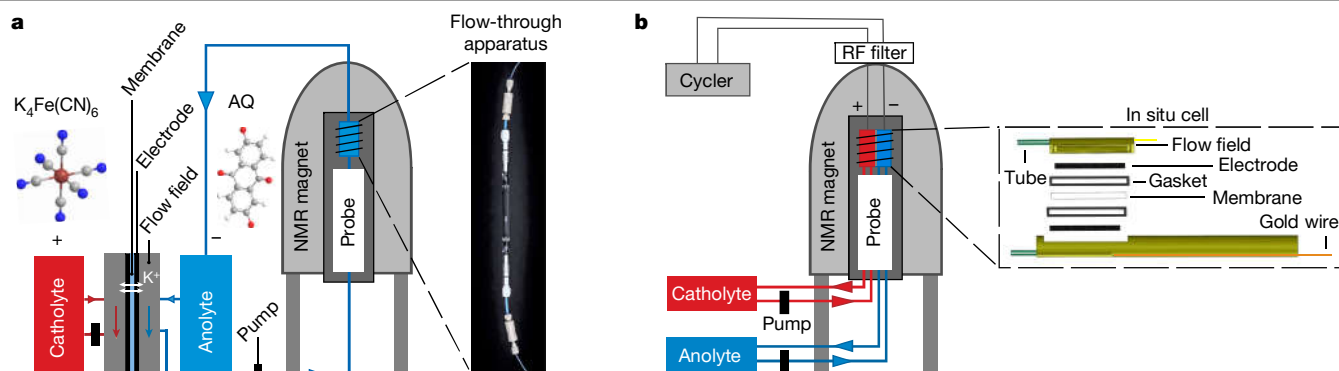


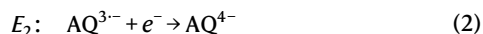
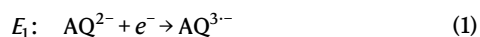
Fig. 1 | Schematics of the two in situ NMR setups. **a**, In the on-line setup, the battery comprises 5.0 cm² carbon felt electrodes, with a catholyte and an anolyte of potassium ferrocyanide(II) and AQ, respectively, dissolved in 1M KOH in D₂O. The volume of the flow path through the magnet including the sampling apparatus (7.3 cm³) and excluding the reservoir is 15.0 cm³. At a flow rate of 13.6 cm³ min⁻¹, the time of flight of the electrolyte out of and back into

the reservoir is 1.1 min. **b**, In the operando setup, the miniaturized flow cell (shown on the right) consists of flow fields, tubes to flow electrolyte in and out, carbon electrodes, a cation-transport membrane and current collectors. The volume of the flow path including the cell cavity (0.032 cm³) is 7.8 cm³. At a flow rate of 2.5 cm³ min⁻¹, the time of flight of the electrolyte out of and back into the reservoir is 3.1 min. RF, radiofrequency.

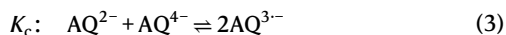
operando setup (0.032 cm³). Furthermore, on-line detection provides superior spectroscopic resolution (Extended Data Fig. 2), because there is no interference from the heterogeneous (particularly the metallic) battery components, which lead to magnetic field inhomogeneities in the operando setup^{11,12}.

Unravelling reaction mechanisms

On-line NMR measurements were performed for a full cell with 20 cm³ of 100 mM anthraquinone (AQ) and 300 mM potassium hexacyanoferrate(II) as the anolyte and catholyte, respectively. On charging at a constant current of 100 mA (20 mA cm⁻²), corresponding to the reduction of the AQs, the battery voltage increases from 1.2 V to the cut-off voltage of 1.7 V for DHAQ (Fig. 2a). Only one voltage step was observed, which is consistent with cyclic voltammetry (CV) that reveals a single reversible redox peak centred at -0.68 V versus the standard hydrogen electrode (SHE; Extended Data Fig. 4a). Despite the single peak, a two-step, single-electron process with half-cell potentials of E_1 and E_2 , has been proposed previously^{5,13}, defined by the following reactions:



A chemical comproportionation reaction then occurs:

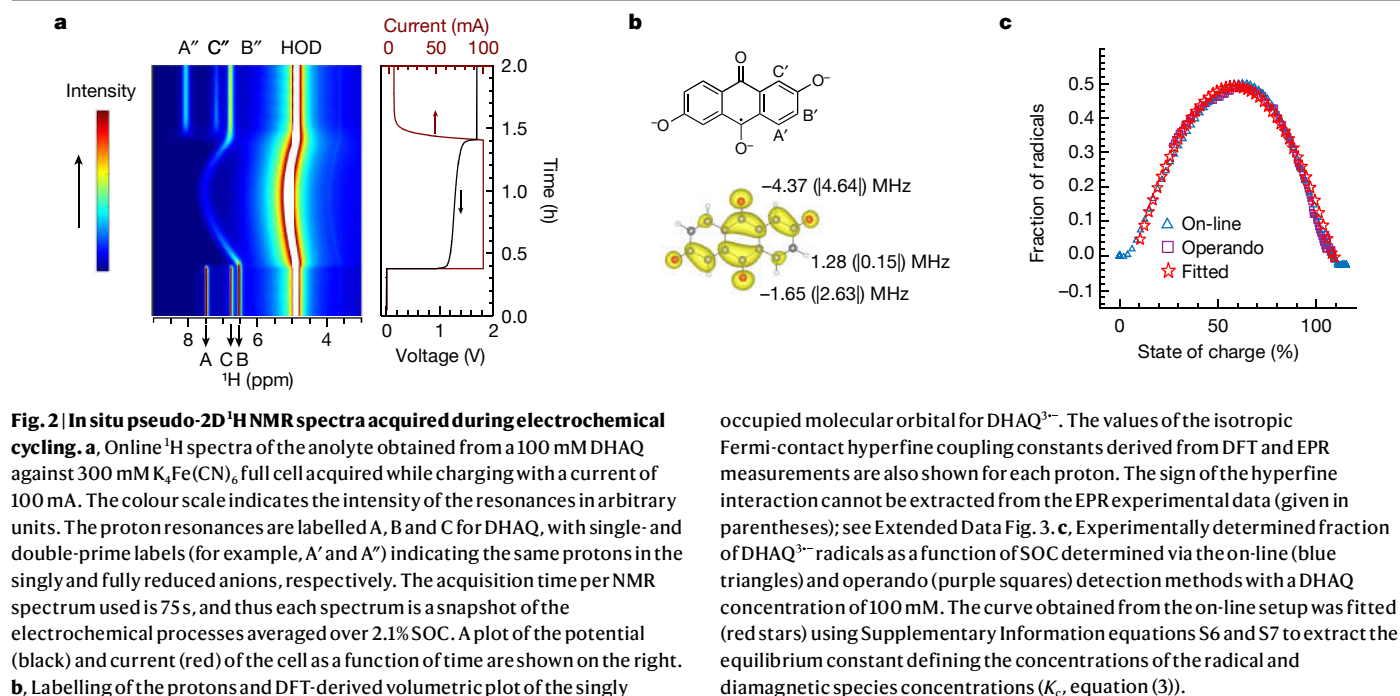


This equilibrium, quantified via the equilibrium constant K_c , controls the concentration of radicals in the solution throughout the electrochemical reactions; however, we note that no direct spectroscopic evidence for radical formation and the complete reduction to AQ⁴⁻ has so far been observed.

Figure 2a presents the ¹H NMR spectra of DHAQ as a function of electrochemical cycling. Upon charging, the proton signals closest to the carbonyl redox centre (A and C) disappear almost immediately, whereas the proton signal farthest from the redox centre (B) broadens and moves towards higher chemical shifts. The apparent loss of signals A and C is ascribed to electron delocalization over the semiquinone radical anion, which results in substantial line broadening^{14,15}. As charging continues, the chemical shift of B reaches a maximum and then moves back towards lower values and narrows as the semiquinones continue to be further reduced. When the cut-off voltage is reached and the potential is held at 1.7 V, proton signals of the final diamagnetic product DHAQ⁴⁻ (A', B' and C') appear. A similar trend was observed

for DBEAQ²⁻, in which the proton signals closest to the carbonyl redox centre disappear almost immediately upon charging, whereas the proton signals farthest from the redox centre move towards a higher chemical shift and then back to lower values until the signals of fully reduced DBEAQ⁴⁻ appear (Extended Data Fig. 2d). Galvanostatic cycling reveals that these changes are reversible (Extended Data Fig. 5).

The broadening of proton resonances is related to the electron delocalization over the radical anion: the higher the electron density on the proton, the broader the peaks. Figure 2b shows the singly occupied molecular orbitals for DHAQ³⁻ (determined by density functional theory (DFT) calculations), and the hyperfine coupling constants determined by EPR at a low concentration (1 mM; Extended Data Fig. 3). The magnitudes of the EPR-derived hyperfine coupling constants corresponding to proton peaks B', A' and C', are |0.15| MHz (B') < |2.63| MHz (A') < |4.64| MHz (C'), which are in agreement with the relative shifts and line broadenings of the corresponding proton resonances. Differences between the hyperfine coupling constants determined by DFT and EPR are ascribed to errors inherent to the DFT method and the lack of inclusion of, for example, solvent effects (Extended Data Fig. 3). Of note, the shift of the water solvent resonance (measured here via the HOD signal present in the predominantly D₂O solvent) mirrors the behaviour of resonance B from DHAQ²⁻ (Fig. 2a), where the shift is ascribed to bulk magnetic susceptibility effects, which are induced by changes in the magnetic susceptibility of the solution. The concentration of radicals can be readily estimated from this bulk magnetic-susceptibility shift using the Evans method (Supplementary Information equations S8–S16), a well established NMR method for measuring the magnetic susceptibility of a solution¹⁶. As the fraction of radicals is directly related to the comproportionation equilibrium constant K_c (equation (3); see also Supplementary Information equations S1–S7) and the state of charge (SOC), the plot of radical concentration versus the SOC can then be fitted to the analytical expressions of Supplementary Information equations S6 and S7 (see Extended Data Fig. 4 for an in-depth explanation) to extract K_c . The fit for DHAQ is shown in Fig. 2c, yielding a K_c of 3.72, corresponding to a potential separation ($E_1 - E_2$) of 33 mV (±10 mV at 293 K; errors are ±half the full-width at half-maximum, FWHM) (Supplementary Information equation S4), in agreement with our CV-model fitting (Extended Data Fig. 4h). Methods section 'Equilibrium concentrations of DHAQ²⁻, DHAQ³⁻ and DHAQ⁴⁻, and CV fittings' discusses assumptions and errors associated with the two approaches for deriving $E_1 - E_2$. Similar results are obtained via on-line and operando detection suggesting that the kinetics of the system is in equilibrium. In the operando experiment, the shift of the water resonances was only quantified (and converted into the radical



concentration) after there was a clear peak separation between the water signals arising from DHAQ and $\text{K}_4\text{Fe}(\text{CN})_6$ electrolytes (Extended Data Fig. 2a).

Determination of the intermolecular electron-transfer rate

The rapid loss of proton signals A and C of DHAQ^{2-} upon formation of fewer than 5% radicals (Fig. 2a) suggests a rapid intermolecular electron-transfer process between the diamagnetic and paramagnetic ions, as described by the bimolecular reaction shown in Fig. 3a. NMR has been previously applied to study such processes^{14,17,18}; suitable approximations have been derived enabling the electron-transfer rate constants to be extracted from the peak broadening (see Supplementary Information and ref. 19). In the slow-exchange regime, the line broadening is proportional to the electron-transfer rate constant k_{ex} (Supplementary Information equation S24), whereas in the fast-exchange regime,

the line broadening is inversely proportional to k_{ex} (Supplementary Information equation S25). As the temperature increases, line broadening of proton signals in the slow-exchange regime should increase, whereas line broadening in the fast-exchange regime should decrease or remain constant.

Variable-temperature ex situ NMR experiments were performed for a 100 mM DHAQ electrolyte solution containing 5 mM DHAQ^{3-} radicals generated by electrochemical reduction (Fig. 3b and Extended Data Fig. 6), in which the radical concentration was estimated by assuming that it is directly proportional to the applied charge. This assumption is valid at the beginning of the reduction/charge, because the concentration of DHAQ^{4-} is small (Extended Data Fig. 4e). As the temperature increases from 283.5 K to 313 K, the linewidths of peaks A and C increase, consistent with slow exchange. The width of peak B remains largely unchanged, suggesting that the exchange is in the fast-exchange

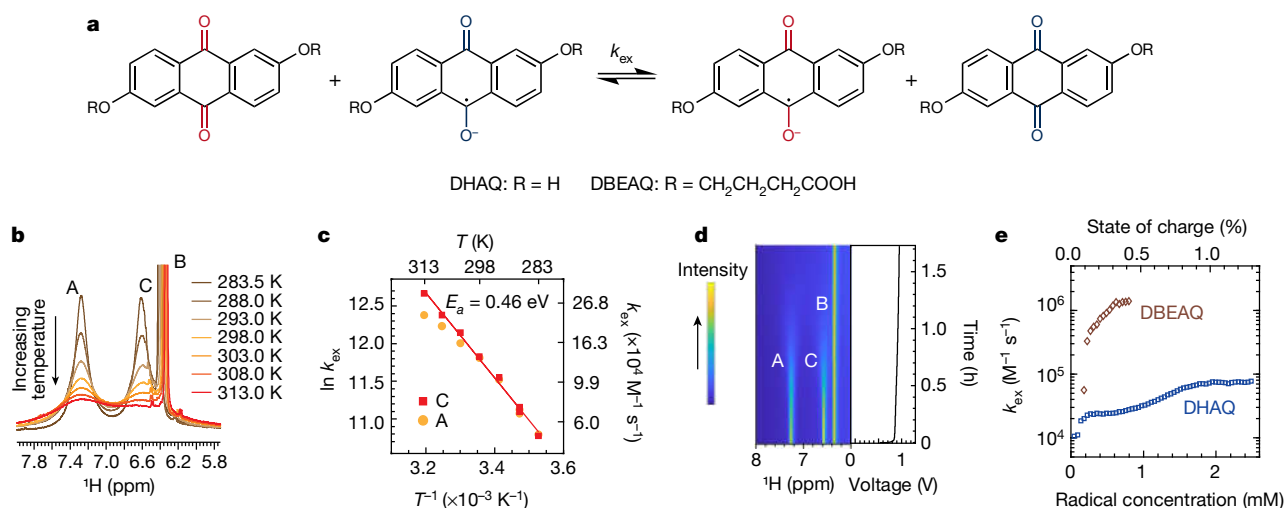


Fig. 3 | NMR analyses of self-exchange electron-transfer reactions. **a**, Self-exchange electron-transfer reactions between oxidized and singly reduced AQs. **b**, Variable-temperature NMR spectra of 5 mM DHAQ^{3-} in a DHAQ solution of total concentration 100 mM. **c**, Arrhenius plot of the electron-transfer rate constant k_{ex} , calculated on the basis of the FWHM of the DHAQ^{2-} peaks A

(coefficient of determination $R^2 = 98.15\%$) and C ($R^2 = 99.49\%$). **d**, In situ ^1H NMR spectra and voltage of the cell comprising 100 mM DHAQ and 300 mM $\text{K}_4\text{Fe}(\text{CN})_6$, obtained with a low current of 1 mA so as to capture the initial stages of charging. **e**, k_{ex} calculated at different radical concentrations and SOC for 100 mM DHAQ and 100 mM DBEAQ.

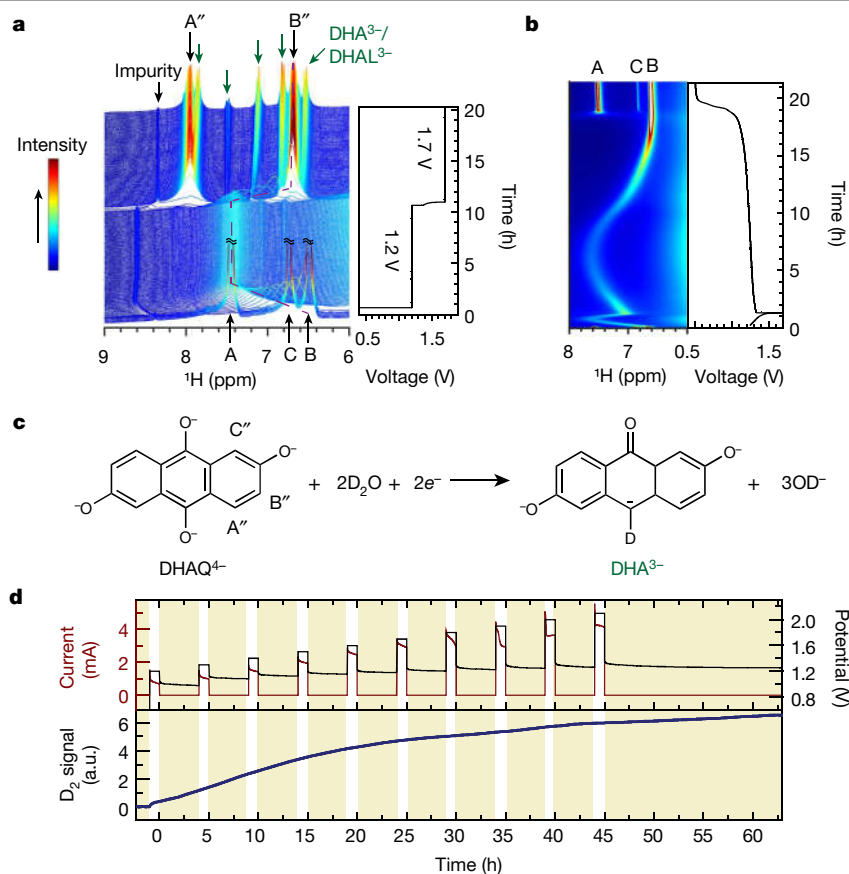


Fig. 4 | In situ NMR and mass spectrometry analyses of electrolyte decomposition and battery self-discharge. a, ^1H NMR spectra of 100 mM DHAQ during a potential hold at 1.2 V and 1.7 V, following charging at 100 mA. Green arrows highlight the DHA^{3-} or DHAl^{3-} signal. The purple dashed line tracks the signal of protons B and B'. The disappearance of signal C' is caused by a H–D exchange reaction with the D_2O solvent (Supplementary Fig. 2). The signal at 8.4 ppm arises from an impurity. **b,** ^1H NMR spectra, voltage and current of a 100 mM 2,6-DHAQ in D_2O in a charge–rest experiment in N_2 (flow)

atmosphere. **c,** Proposed decomposition reaction of DHAQ^{4-} . **d,** In situ mass spectrometry signals of D_2 (mass-to-charge ratio, $m/z = 4$) during a stepped potential-hold experiment (black, potential; red, current). The potential was increased from 1.2 V to 2.1 V, stepping in increments of 0.1 V, holding for 1 h at each potential step with an interleaved 4 h rest period. 8 cm^3 of 100 mM DHAQ and 8 cm^3 of 300 mM $\text{K}_4\text{Fe}(\text{CN})_6$ dissolved in D_2O are loaded in each electrolyte compartment of a custom-made H cell. a.u., arbitrary units.

regime¹⁹ (see Methods). The FWHM of peaks A and C were then used to calculate the electron-transfer rate constant k_{ex} under the slow-exchange approximation, in which the line broadening—strictly, the transverse relaxation rate—caused by exchange with the paramagnetic ions is given by the expression $R_{2p} = k_{\text{ex}}[P]$, where $[P]$ is the radical concentration. As shown in Fig. 3c, the plot of $\ln k_{\text{ex}}$ versus temperature, as calculated from peak C, is linear, whereas that for peak A deviates from linear behaviour at elevated temperatures as proton A'—with its smaller electron spin density—approaches the intermediate-exchange regime¹⁹. An activation energy of 0.46 eV for the electron-transfer reaction between DHAQ^{2-} and DHAQ^{3-} is obtained, a value that is larger than, for example, the DFT-derived values of 0.28 eV and 0.34 eV for electron transfer between sulfonated AQs²⁰.

After validating the applicability of the slow-exchange approximation, the self-exchange electron-transfer processes are readily investigated in situ at room temperature (293 K; Fig. 3d and Extended Data Fig. 7). Electron-transfer rate constants k_{ex} of approximately $1 \times 10^5\text{ M}^{-1}\text{ s}^{-1}$ and $1 \times 10^6\text{ M}^{-1}\text{ s}^{-1}$ for DHAQ and DBEAQ, respectively—as shown in Fig. 3e and Extended Data Fig. 7i—were extracted from the broadening of peaks A and C of DHAQ and D, E and F of DBEAQ. The changes in the calculated rate constants k_{ex} at the onset of reduction are ascribed to the inhomogeneous mixing of the low AQ^{3-} fraction (<1%) with AQ^{2-} as the electrolytes leave the reactor and flow through the tubing into the NMR magnet. Constant (equilibrium) k_{ex} values are obtained when the reaction progresses, because the radical anions

become homogeneously distributed, evidenced by the plateau at higher radical concentrations. The higher intermolecular electron-transfer rate for DBEAQ is probably due to (1) the weaker intermolecular Coulombic repulsions between DBEAQ anions than between DHAQ anions, because the negative charges are more dispersed on DBEAQ than on DHAQ, and (2) the stronger van der Waals and hydrophobic interactions between the longer (non-polar) *R* groups in DBEAQ. The values of DHAQ and DBEAQ based on the in situ and ex situ NMR analysis are in the range of the previously reported values^{20–23} on other organic systems, spanning $10^4\text{ M}^{-1}\text{ s}^{-1}$ to $10^{10}\text{ M}^{-1}\text{ s}^{-1}$.

Following electrolyte decomposition and battery self-discharge in real time

The in situ NMR approach enables us to follow electrolyte decomposition under specific cycling conditions. For example, for DHAQ, new ^1H NMR signals were observed at 6.5, 6.8, 7.1, 7.7 and 7.9 ppm during a potential hold at 1.7 V, performed to ensure the complete reduction of DHAQ (Fig. 4a). On the basis of ex situ two-dimensional (2D) NMR correlation experiments and DFT-derived chemical shifts (Extended Data Fig. 8), these signals are assigned to the degradation products anthrone (DHA^{3-}) and anthrol (DHAl^{3-}), as identified previously²⁴. Subsequent galvanostatic cycling of the solution without a potential-hold step did not substantially change the signal intensity of DHA^{3-} and DHAl^{3-} (Extended Data Fig. 8e). By contrast, when the potential hold was reduced to 1.2 V, a much smaller concentration of decomposition

products was detected (peaks at 6.7, 7.1 and 8.1 ppm; Fig. 4a), suggesting that the nature and extent of decomposition depends on the reduction potential. These observations suggest that the decomposition is an electrochemical reaction with a possible route outlined in Fig. 4c. A chemical disproportionation reaction with water, as previously proposed²⁴, cannot be excluded; however, the voltage dependence suggests that the products seen here are formed electrochemically. By contrast, for DBEAQ, no decomposition products were detected during a potential hold at 1.4 V and 1.7 V (Extended Data Fig. 9), which is in agreement both with previous work¹³ and with the proposal that degradation of DBEAQ occurs over long-term cycling via a mechanism involving nucleophilic attack, rather than by degradation of the reduced species¹³.

The in situ NMR technique can be readily used to monitor battery self-discharge. As shown in Fig. 4b, the open-circuit voltage of a charged battery (under the protection of flowing N₂ gas) slowly decreases from 1.3 V to 1.1 V, and then after 17 h a sharp decrease to 0.5 V is seen. The in situ NMR spectra show that this is caused by the reoxidation of DHAQ⁴⁻ to DHAQ³⁻ and DHAQ²⁻. The rapid drop in open-circuit voltage occurs when the DHAQ²⁻ signals A and B reappear and sharpen, indicating that the drop is due to complete oxidation of the solution. To identify the potential oxidant, we developed an in situ mass-spectrometry technique based on the design of the H cell to monitor the gas evolution (Fig. 4d). A stepped-potential experiment was performed from 1.2 V to 2.1 V, in increments of 0.1 V. D₂ evolution commences at 1.2 V, and it is steadily evolved both at higher potentials and during the rest periods. These two observations suggest that D₂ evolution originates from a chemical reaction, most probably from water reduction: DHAQ⁴⁻ + 2D₂O → DHAQ²⁻ + D₂ + 2OD⁻. We note that the redox potential of DHAQ is higher than that of the hydrogen evolution from water (-0.83 V versus SHE)¹³ in a 1 M KOH aqueous solution, making this reaction thermodynamically unfavourable (unless substantial pH fluctuations occur). We cannot rule out the further reaction of the degraded electrolytes, because there is considerable literature precedent for reactions of AQ-based structures that involve hydrogen evolution—anthrone dimerization, for example, forms bianthrone accompanied by hydrogen evolution²⁵. In addition, some reported systems are photosensitive, providing another potential degradation route²⁶. Further investigation into the various degradation mechanisms is in progress to understand how to control the stability of anthrahydroquinones in aqueous media.

In summary, we have demonstrated two in situ NMR metrologies to study flow batteries. The formation of radicals and fully reduced anions is directly observed in two AQ-based redox flow battery systems, in which their equilibrium concentrations are governed by the potentials of the two single-electron-transfer redox processes. The radical concentrations as a function of SOC were quantified by analysing the bulk magnetic susceptibility changes, enabling the voltage separation of the two successive reductions to be extracted. The redox reaction was found to be coupled with electron transfer between the radicals and diamagnetic anions, with NMR spectroscopy providing a method to measure the rates of these reactions. The presence of self-exchange electron-transfer reactions in organic flow batteries has major implications because it affects the overall rates of the redox reactions, controlling, for example, the disproportionation rate. Electrochemically triggered decomposition of DHAQ⁴⁻ to DHA³⁻/DHAL³⁻ was observed under specific cycling conditions, but no decomposition of DBEAQ⁴⁻ was observed. The real-time observation of reoxidation of DHAQ⁴⁻ and hydrogen evolution in these aqueous media indicates that other side reactions are occurring that involve solvent water and/or degradation of DHAQ⁴⁻. Owing to the simplicity of the on-line NMR setup—which consists, in essence, of a laboratory-scale redox flow battery and a flow NMR sampling tube—we expect wide adoption of this technique, which will help to advance the understanding of various redox chemistries in flow batteries and other electrochemical systems. Beyond battery

research, we demonstrate a way to study radical species, particularly at high radical concentrations when hyperfine coupling features in an EPR spectrum are lost owing to both electron-spin interactions and electron-transfer reactions. Our work shows that by following spectral changes in real time, NMR can provide key information concerning molecular structure, spin-density distributions and intermolecular electron-hopping rates.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2081-7>.

- US Energy Information Administration. *Annual Energy Outlook 2018* (US Department of Energy, 2018).
- Wei, X. et al. Materials and systems for organic redox flow batteries: status and challenges. *ACS Energy Lett.* **2**, 2187–2204 (2017).
- Leung, P. et al. Recent developments in organic redox flow batteries: a critical review. *J. Power Sources* **360**, 243–283 (2017).
- Yang, B., Hooper-Burkhardt, L., Wang, F., Prakash, G. K. S. & Narayanan, S. R. An inexpensive aqueous flow battery for large-scale electrical energy storage based on water-soluble organic redox couples. *J. Electrochem. Soc.* **161**, A1371–A1380 (2014).
- Lin, K. et al. Alkaline quinone flow battery. *Science* **349**, 1529–1532 (2015).
- Tong, L. et al. UV-vis spectrophotometry of quinone flow battery electrolyte for in situ monitoring and improved electrochemical modeling of potential and quinhydrone formation. *Phys. Chem. Chem. Phys.* **19**, 31684–31691 (2017).
- Lawton, J. S., Aaron, D. S., Tang, Z. & Zawodzinski, T. A. Electron spin resonance investigation of the effects of vanadium ions in ion exchange membranes for uses in vanadium redox flow batteries. *ECS Trans.* **41**, 53–56 (2012).
- Webster, R. D. In situ electrochemical-NMR spectroscopy. Reduction of aromatic halides. *Anal. Chem.* **76**, 1603–1610 (2004).
- Friedl, J. et al. Asymmetric polyoxometalate electrolytes for advanced redox flow batteries. *Energy Environ. Sci.* **11**, 3010–3018 (2018).
- Cao, S.-H. et al. In situ monitoring potential-dependent electrochemical process by liquid NMR spectroelectrochemical determination: a proof-of-concept study. *Anal. Chem.* **89**, 3810–3813 (2017).
- Pecher, O., Carretero-González, J., Griffith, K. J. & Grey, C. P. Materials' methods: NMR in battery research. *Chem. Mater.* **29**, 213–242 (2017).
- Wang, H. et al. In situ NMR spectroscopy of supercapacitors: insight into the charge storage mechanism. *J. Am. Chem. Soc.* **135**, 18968–18980 (2013).
- Kwabi, D. G. et al. Alkaline quinone flow battery with long lifetime at pH 12. *Joule* **2**, 1894–1906 (2018); correction **2**, 1907–1908 (2018).
- de Boer, E. & MacLean, C. NMR study of electron transfer rates and spin densities in *p*-xylene and *p*-diethylbenzene anions. *J. Chem. Phys.* **44**, 1334–1342 (1966); erratum **47**, 3102 (1967).
- Pell, A. J., Pintacuda, G. & Grey, C. P. Paramagnetic NMR in solution and the solid state. *Prog. Nucl. Magn. Reson. Spectrosc.* **11**, 1–271 (2019).
- Evans, D. F. The determination of the paramagnetic susceptibility of substances in solution by nuclear magnetic resonance. *J. Chem. Soc.* **1959**, 2003–2005 (1959).
- Johnson, C. S. Jr. Nuclear transverse relaxation in electron-transfer reactions. *J. Chem. Phys.* **39**, 2111–2114 (1963).
- de Boer, E. & MacLean, C. Spin densities in the alkyl groups of alkyl-substituted naphthalene negative ions, determined by NMR. *Mol. Phys.* **9**, 191–193 (1965).
- Bertini, I., Luchinat, C., Parigi, G. & Ravera, E. *NMR of Paramagnetic Molecules* 2nd edn (Elsevier, 2017).
- Rosso, K. M., Smith, D. M. A., Wang, Z., Ainsworth, C. C. & Fredrickson, J. K. Self-exchange electron transfer kinetics and reduction potentials for anthraquinone disulfonate. *J. Phys. Chem. A* **108**, 3292–3303 (2004).
- Grampp, G., Galán, M. & Sacher, M. Kinetics of photoinduced electron transfer reactions of some anthraquinone radical anions with various inorganic ions: comparison with Marcus cross-relation. *Ber. Bunsenges. Phys. Chem.* **99**, 111–117 (1995).
- Rosokha, S. V. & Kochi, J. K. Continuum of outer- and inner-sphere mechanisms for organic electron transfer. Steric modulation of the precursor complex in paramagnetic (ion-radical) self-exchanges. *J. Am. Chem. Soc.* **129**, 3683–3697 (2007).
- Meisel, D. & Fessenden, R. W. Electron exchange and electron transfer of semiquinones in aqueous solutions. *J. Am. Chem. Soc.* **98**, 7505–7510 (1976).
- Goulet, M.-A. et al. Extending the lifetime of organic flow batteries via redox state management. *J. Am. Chem. Soc.* **141**, 8014–8019 (2019).
- Donovan, P. M. & Scott, L. T. 4,11-bisanthrenequinone and 10,10'-bianthrone: simple one-step syntheses from anthrone. *Polycycl. Aromat. Compd.* **28**, 128–135 (2008).
- Filipescu, N., Avram, E. & Welk, K. D. Photooxidative transformations of anthrone, bianthrone, and bianthrone in acid solution. *J. Org. Chem.* **42**, 507–512 (1977).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Materials and synthesis

2,6-dihydroxyanthraquinone (2,6-DHAQ, A89502 technical grade, 90% purity), potassium hexacyanoferrate(II) trihydrate (P3289, $\geq 98.5\%$ purity), D_2O (151882, 99.9 atom%) and ethyl 4-bromobutyrate (167118, 95% purity) were purchased from Sigma Aldrich Chemicals. Anhydrous *N,N*-dimethylformamide (43465, $\geq 99.9\%$ purity), potassium ethoxide (14263.30, $\geq 95\%$ purity), anhydrous potassium carbonate (A16625, $\geq 99.0\%$ purity), isopropanol (20839.366, $\geq 99.0\%$ purity) and glacial acetic acid (20103.364, 99–100%) were purchased from VWR.

4,4'-((9,10-anthraquinone-2,6-diyl)dioxy)dibutyric acid (2,6-DBEAQ) was synthesized following a previously reported method¹³. 2,6-DHAQ (5.00 g, 20.8 mmol) was dissolved in anhydrous DMF (250 cm³) and potassium ethoxide (6.13 g, 72.8 mmol) added to the solution under vigorous stirring. The mixture was stirred at room temperature for 15 min. To this, anhydrous K_2CO_3 (14.40 g, 104.2 mmol) and ethyl 4-bromobutyrate (21.41 g, 109.8 mmol) were added and the mixture solution was then heated to 95 °C for 23 h. The reaction mixture was poured into deionized water, and the solid collected by filtration, washed with deionized water and dried in a vacuum oven overnight to afford diethyl 4,4'-((9,10-anthraquinone-2,6-diyl)dioxy)dibutyrate (2,6-DEDBEAQ, 9.20 g, 94.2%), the ester precursor of 2,6-DBEAQ, as silver-grey powders. 2,6-DEDBEAQ (1.00 g, 2.13 mmol) was hydrolysed in a 0.15 M KOH water:isopropanol (2:1 v/v) solution (60 cm³) at 60 °C for 12 h. The reaction mixture was poured into deionized water (200 cm³) and glacial acetic acid added dropwise to adjust the pH to 4. The solid was collected by filtration, washed with deionized water and dried in a vacuum oven overnight to afford 2,6-DBEAQ in a quantitative yield. The NMR spectra that were subsequently acquired were consistent with previously reported spectra¹³.

Flow battery assembly

The hardware of the flow battery was purchased from Scribner Associates. Ultrahigh-purity sealed graphite flow plates with serpentine flow patterns were used for both electrodes. Each electrode comprised 4.6 mm or 6.0 mm carbon felt (SGL) with a 5 cm² active area. Nafion 212 was used as the ion transport membranes. Pretreatment of the Nafion 212 membranes was performed by first heating the membrane in 80 °C deionized water for 20 min and then soaking it in 5% hydrogen peroxide solution for 35 min. The treated membranes were stored in 0.1 M KOH solution at room temperature. PTFE frames with a thickness of 3 mm were used to position the electrodes with Viton gaskets 0.7 mm in thickness on each side of the frames. The current collectors were gold-plated copper plates. Anodized aluminium end plates with reactant input/output ports were used; however, it was found that the coatings were easily corroded by basic solutions. Direct solution contact was avoided by carefully inserting the inlet and outlet tubing through the Viton O-ring seals.

A stock solution of 1 M KOH dissolved in D_2O was prepared and used as the solvent. On the negative side, either 200 mM (in 1.4 M KOH), 100 mM, 50 mM or 30 mM 2,6-DHAQ was prepared in 20 cm³ of solvent. The SOC of the battery was calculated by dividing the number of electrons that have flowed from the electrochemical cycle by the theoretical storage capacity of the quinone anions, assuming a two-electron-per-anion redox process. For example, for a 100 mM DHAQ of 90% purity, the theoretical capacity is 96.5 mAh. On the positive side, potassium hexacyanoferrate(II) trihydrate was dissolved in 20 cm³ of solvent to form a 300 mM solution; this corresponds to 1.5 times the total capacity of the 100 mM 2,6-DHAQ solution for the same volume of electrolyte. When the battery undergoes galvanostatic cycling, 50 mM potassium hexacyanoferrate(III) is added to the positive side to ensure that both Fe^{2+} and Fe^{3+} are in excess during battery cycling.

Custom-made glassware made from Pyrex with gas inlet, outlet, liquid inlet and outlet were used as electrolyte reservoirs. Prior to the

experiments, all solutions were degassed with N_2 gas rigorously for 30 min to 1 h. The torque applied on the bolts that tighten the cell was found to affect the battery performance. It was optimized at 2 N m on each bolt. The galvanostatic cycling of the battery was controlled by a portable potentiostat (SP-150, BioLogic SAS).

On-line NMR setup

A custom-made medium-wall flow-through NMR sampling tube of 14 cm in length and 10 mm outer diameter was positioned in a micro-imaging probe (Bruker 2.5; Extended Data Fig. 1a). The electrolyte solution flows from the bottom to the top of the tube. The inlet and outlet of the sampling tube were connected to two 1/16 inch PFA tubes (0.5 mm inner diameter) 3 m in length via 1/16 inch to 1/8 inch tube adaptors. The PFA tube at the bottom is connected to the outlet of the battery; the PFA tube at the top is connected to the inlet of the electrolyte reservoir. The electrolyte is pumped through the sampling tube and the flow battery which is positioned next to a 300 MHz NMR magnet outside the 5 gauss (G) line by a peristaltic pump (MasterFlex L/S 07551-20, Cole-Parmer; pump head, MasterFlex tubing 77202-60 #14, Chem Bio). The volume of the electrolyte inside the NMR sampling tube is 7.3 cm³. The volume of the electrolyte in the PFA tubes is 1.2 cm³. The volume of the electrolyte that passes through the MasterFlex tube is 6.5 cm³. On the basis of an inner diameter of 8.16 mm of the sampling tube and 3 cm of the detection length by the NMR probe, the effective volume detected by NMR is 1.57 cm³. Flow rates were measured at different rotary speeds of the pump using the same tubing, as shown in Extended Data Fig. 1d. At a flow rate of 13.6 cm³ min⁻¹, the electrolyte takes 60 s to flow back to the reservoir, so the time lag between the electrochemical cycling and the NMR detection is 30 s.

As shown in Extended Data Fig. 1c, pseudo-2D NMR experiments were performed by direct excitation with a 90° pulse. The acquisition time is 1.5 s. Each spectrum has a number of scans, $n_s = 8$. After eight scans, a time delay d_2 is introduced before the next spectrum acquisition starts. With a recycle delay, $d_1 = 7$ s, and $d_2 = 7$ s, the total acquisition time of a spectrum is 75 s ((7 + 1.5) × 8 + 7 = 75 s). Owing to the high ionic conductivity of the basic solution, the 90° pulse width for a proton increases from 20 μs for a non-conductive solution to 27 μs at a radiofrequency power of 30 W. Pulse calibration was performed as a function of SOC and the 90° pulse width remained the same. All spectra were referenced to the water chemical shift at 4.79 ppm before battery cycling started. The spectral widths were 200 or 20 ppm. Peak assignment was facilitated by the scalar coupling (J) between protons A and B, J_{AB} , and between protons B and C, J_{BC} . The J -coupling interactions among the three aromatic protons were $H(J_{AB}) = 8.6$ Hz and $H(J_{BC}) = 2.5$ Hz.

Operando NMR setup

The in situ cell assembly consists of six key components: the PEEK flow field, carbon electrode, an ion-transport membrane, current collectors, a PEEK sleeve and PEEK tubes (flow inlet and outlet). Pictures of the in situ cell are shown in Extended Data Fig. 1b. The diameter of the cell assembly is 9.9 mm, which fits into a micro-imaging probe (Bruker 2.5). The structures of the catholyte and anolyte compartment are identical except for an extended solid part 12.20 cm in length attached to one compartment, which is held in the NMR probe by a screw cap. The flow field inside the electrolyte compartment is 1.80 cm in length, 0.40 cm in width and 500 μm in depth. A layer of carbon electrode (39AA, 80% porosity, Sigracet) of 1.80 cm in length, 0.40 cm in width and 280 μm in thickness is placed inside the flow field. A current collector made of two gold wires of outer diameter 0.5 mm and length 16.0 cm passes through a hole at the back of the flow field and is in electrical contact with the carbon electrode. The other end of the current collector is connected to an electrical cable that passes through a d.c. 5 MHz low-pass filter at the top of the NMR magnet and connects to the electrochemical cycler. A treated Nafion 212 membrane 2.80 cm in length, 0.80 cm in width and 50.8 μm in thickness is compressed between the catholyte

and anolyte compartment by Viton O-rings. The cell assembly is held in a PEEK sleeve of outer diameter 0.99 cm and inner diameter 0.89 cm. The electrolyte inlet and outlet on each compartment are PEEK tubes (outer diameter, 1/16 inch) 10 cm in length. They are connected to 1/16 inch PFA tubes 3 m in length via 1/16 inch to 1/16 inch unions that pass through the bottom of the NMR probe. The PFA tube is connected to a peristaltic pump (MasterFlex L/S 07551-20, Cole-Parmer; pump head, MasterFlex tubing 77202-60 #14, Chem Bio), and then to the electrolyte reservoir. The total volume of electrolyte in the PFA tubes is 1.2 cm³. The volume in the MasterFlex tube on the pump is 6.5 cm³. At a flow rate of 2.5 cm³ min⁻¹, the electrolyte takes 185 s to flow back to the reservoir, and so the time lag between the electrochemical cycling and the NMR detection is 92 s. The effective NMR detection volume in the in situ cell is 0.032 cm³ (excluding the volume of carbon electrode), and therefore the residence time in the detection region of the NMR probe is 0.8 s. Pseudo-2D NMR experiments were performed with a 90° pulse of 27 μs, an acquisition time of 0.15 s ($d_1 = 1$ s, $d_2 = 1$ s, $n_s = 64$). The total acquisition time of a spectrum is 75 s ($(1 + 0.15) \times 64 + 1 = 74.6$ s).

In situ relaxation measurements

The longitudinal (T_1) and transverse (T_2) relaxation time measurements were performed in situ via the on-line setup. A full battery with 20 mL of 100 mM DHAQ and 40 mL of 200 mM K₄Fe(CN)₆ was charged at 100 mA to the cutoff potential of 1.7 V. The potential was then held at 1.7 V for 40 min. A flow rate of 13.6 cm³ min⁻¹ was used. The flow and electrochemical cycling were paused intermittently during charging (every 5 or 10 min, see Extended Data Fig. 10), and the relaxation measurements were performed on a static solution. The 90° pulse width was calibrated before each measurement and was found to be the same throughout the electrochemical cycling. After each relaxation measurement, flow and electrochemical cycling were resumed. The relaxation measurement at each SOC takes up to 30 min, which is much shorter than the time for the reoxidation of DHAQ (17 h). There was no noticeable change of chemical shift before and after each measurement, suggesting that the reduced DHAQ is stable on the time scale of the relaxation measurement. Nonetheless, there will be errors arising from the reoxidation and decomposition of DHAQ during the measurement, particularly at high SOC (as discussed in the main text). The errors are reflected in the data fitting and are shown in Extended Data Fig. 10b, c.

T_1 relaxation time measurements were performed with an inversion-recovery (tlir) pulse sequence comprising 180° and a 90° pulses, with an increasing delay time (t) between the two pulses for each spectrum in the second dimension. Eight to 16 spectra were acquired in the second dimension. A recycle delay of 30 s was applied for the oxidized DHAQ. As the T_1 time of proton B in DHAQ in the presence of radicals is an order of magnitude shorter than that of HOD, two separate measurements were performed on DHAQ proton B and HOD, with a d_1 of 0.5 s and 2 s used for each molecule. The signal integral $I(t)$ was plotted as a function of t and fitted by $I(t) = I_0(1 - 2e^{-t/T_1})$, where I_0 is the maximum peak integral.

T_2 relaxation time measurements were performed with a Carr–Purcell–Meiboom–Gill pulse sequence comprising a 90° pulse and a train of 180° pulses, with a delay time of 1 ms before and after each 180° pulse. Eight to 16 spectra were acquired in the second dimension, with an increasing number (up to 2,000) of 180° pulses. The signal integral $I(t)$ was plotted as a function of the sum of the delay time t and fitted with $I(t) = I_0e^{-t/T_2}$.

T_1 relaxation times for the H^A, H^B and H^C protons (corresponding to A, B and C, respectively) and HOD were measured to be 1.6 ± 0.2 , 1.9 ± 0.2 , 4.9 ± 0.2 and 13 ± 1 s, respectively. We note that these T_1 times depend on the protonation level of the deuterated solvent, because the relaxation is largely driven by the proton–dipolar coupling interaction for the diamagnetic solutions. Shorter T_1 relaxation values for H^A, H^B and H^C on the fully reduced DHAQ^{•-} anion were measured to be 0.32 ± 0.06 , 0.34 ± 0.07 , and 0.22 ± 0.08 s, respectively, but these

were measured in the presence of radicals, that is, in the in situ experiments. The errors here represent the 95% confidence level from the fit. T_1 and T_2 times were then measured as a function of SOC. Extended Data Fig. 10a presents the voltage profile of a full battery of 100 mM DHAQ and 200 mM K₄Fe(CN)₆ during intermittent charging with a current of 100 mA. Extended Data Fig. 10b shows that the measured T_1 and T_2 values for HOD both decrease rapidly on charging to an SOC of 9%: T_1 decreasing from 13.1 s to below 0.7 s, and T_2 decreasing from 5.1 s to below 0.5 s. The changes are more gradual thereafter, reaching a minimum of 0.1 s (for both T_1 and T_2) at 50–70% SOC. Both T_1 and T_2 then increase to 0.4 s as the battery is charged to its full capacity. The T_1 value for B follows the same trend as that of HOD, decreasing from 1.9 s to below 0.03 s at an SOC of 9% (Extended Data Fig. 10c).

Effect of flow rate, radical concentration and relaxation times on magnetization build-up and line broadening

Under flow conditions, the build-up of magnetization of the nuclear spins is determined by the time that the electrolyte molecules spend in the high magnetic field²⁷. Since $5T_1$ enables a build-up of 99.3% of the maximum thermal polarization, ideally a residence time of the electrolyte in the field, τ , should be longer than $5T_1$. τ is related to the flow rate v by $\tau = V/v$, where V is the volume of electrolyte in the high magnetic field. Setting V to be the same as the detection volume of the NMR probe gives the lower limit of the residence time, because the region of high magnetic field extends to a longer length than the detection region of the NMR probe. In the on-line setup, the volume of the detection region is 1.57 cm³. A residence time of $5T_1$ of the DHAQ protons H^A ($T_1 = 1.5$ s), H^B (1.9 s), H^C (4.9 s) and HOD (13 s) requires flow rates lower than 12.6, 9.9, 3.8 and 1.4 cm³ min⁻¹, respectively, for quantitative measurements of the diamagnetic species in the absence of any radicals. In the operando cell, the detection volume is 0.032 cm³ and a residence time of $5T_1$ for H^A, H^B, H^C and HOD gives flow rates of 0.3, 0.2, 0.1 and 0.03 cm³ min⁻¹, respectively. We note that if faster flow rates are required, simple methods for polarizing the nuclear spins before the liquids actually enter the radiofrequency coil (for example, by adding loops or liquid reservoirs in the magnet) can be readily added to the setup.

Extended Data Fig. 10d, e shows the ¹H NMR spectra of DHAQ and HOD as a function of flow rate without electrochemically cycling the battery in the on-line setup. The signal integral is plotted against the flow rate, as shown in Extended Data Fig. 10f, g. As the flow rate increases, the decrease of the water signal is the most pronounced, owing to its large value of T_1 , followed by proton C of DHAQ. The signal intensity of A and B is almost unaffected by the flow rates studied here, owing to the shorter T_1 values of H^A and H^B. Of note, when radical species (for example DHAQ^{•-}) are generated, T_1 will be decreased substantially, owing to the nuclei–electron spin interactions, and much higher flow rates will be possible without reducing the signal intensity.

To achieve optimized electrochemical performance of the battery system—that is, a low overpotential—a high flow rate is desirable in order to drive the system out of the mass-transport-limited regime. Given the detection volumes of 1.57 cm³ and 0.032 cm³ in the on-line and operando setups, respectively, flow rates that allow quantitative measurements of DHAQ proton B in the presence of small concentrations of radicals and allow a residence time of $>5T_1$ (5×0.03 s) correspond to <628.0 cm³ min⁻¹ and <14.4 cm³ min⁻¹, respectively. The flow rates in the majority of the on-line and operando NMR experiments are set to 13.6 cm³ min⁻¹ and 2.5 cm³ min⁻¹, respectively, which are in this quantitative regime.

During NMR data acquisition, the thermal polarization is converted into transverse magnetization by a 90° radiofrequency pulse, and the decay of the transverse magnetization (the T_2 relaxation time) determines the linewidth of the NMR signal. There is a finite probability that a fraction of the spins with transverse magnetization will leave the detection region before the data acquisition finishes, which may cause

further line broadening. The longer the T_2 relaxation time is, the more susceptible the signal is to line broadening from flow. The FWHM of the DHAQ and the HOD signals as a function of flow rate are plotted in Extended Data Fig. 10h. As the flow rate increases from 13.6 cm³ min⁻¹ to 37.5 cm³ min⁻¹, the HOD FWHM increases slightly, from 6 to 7 Hz, whereas the FWHM of the DHAQ signals show negligible change. In the presence of the DHAQ^{3•-} radicals, which are produced during battery cycling, T_2 is decreased and thus the effect of flow on the linewidth of the DHAQ proton signals is further reduced.

Study of H–D exchange by ex situ NMR

A solution of 100 mM 2,6-DHAQ dissolved in 1 M KOH in D₂O was prepared and left in an ambient environment for 30 days before taking the first NMR spectrum. The same solution was then charged (corresponding to the electrochemical reduction of 2,6-DHAQ) in a redox flow battery and an aliquot of 0.5 cm³ was extracted. The aliquot was transferred to a 5 mm thin-walled NMR tube and one-dimensional NMR spectra were acquired on a 500 MHz solution NMR spectrometer. For ¹H NMR, 32 scans were accumulated with a 30° pulse and $d_1 = 1$ s. For ¹³C NMR, 1,024 scans were accumulated with a 30° pulse and $d_1 = 3$ s.

Study of electron-transfer reactions by variable-temperature ¹H NMR experiments

The solvent of 1 M KOH in D₂O was degassed by argon gas for 2 h. For the charge–rest experiments, a H cell was assembled inside the glovebox (described below). 20 cm³ of 100 mM DHAQ solution was first reduced at 10 mA for 30 min (9 mM concentration of radicals) and an aliquot of 0.1 cm³ was extracted and sealed in a 5 mm thick-walled NMR tube with an airtight Young's tap. Then the DHAQ was oxidized at 1 mA for 140 min (5 mM of radicals) and an NMR sample was prepared in the same way as for the 9 mM radical concentration.

NMR spectra were acquired on a 500 MHz solution NMR spectrometer using a one-pulse (90°) sequence. For the spectra acquisitions of the sample containing 5 mM of radicals, the temperature of the NMR probe was ramped up—from 283 K to 313 K in increments of 5 K—and then cooled down to 288 K; spectra were acquired at each temperature. The heating was carried out in this way to ensure that any reoxidation of the solution did not perturb the measurement. For the spectral acquisition of the sample containing 9 mM of radicals, the temperature was set at 283 K and then ramped up from 288 K to 338 K in increments of 10 K. Because the magnetic field was locked on the water resonance at 4.79 ppm and the water resonance is temperature-dependent, the shift of peak A of DHAQ was manually set at 7.3 ppm after acquisition to allow for ready comparison of the different spectra.

As shown in the EPR results (Extended Data Fig. 3), the magnitudes of the hyperfine coupling of the three protons are 0.15 MHz (B') < 2.63 MHz (A') < 4.64 MHz (C'). Owing to the much smaller hyperfine coupling constant, the exchanges between the protons that give rise to resonances B and B', and those between B' and B'' are in the fast-exchange regime. (We shall refer to the coalesced resonances as B/B' and B'/B''). By contrast, the exchanges involving A/A' (and A'/A'') and C/C' (and C'/C'') are in the slow-exchange regime. This is also shown by the variable-temperature NMR experiments shown in Extended Data Fig. 6b, c where the linewidth of resonance B/B' is largely insensitive to the changes in the temperature and in the electron-transfer rate.

To verify the dependence of linewidth on the exchange rate k_{ex} , the effect of two-site chemical exchange on the spectra was simulated with the program Spinach²⁸. In the simulation, the concentration of radicals was set to either 5 mM or 9 mM, that is, the same as the radical concentration used in the variable-temperature experiments. The position of resonance B' of 7.51 ppm was used, which is estimated from the shift of B in the on-line NMR spectra (Fig. 2a) at low radical concentrations. For example, in the presence of 1.35% radicals (as determined by the shift of the water resonance via the Evans method), the shift of B is 6.46 ppm, and the shift of B' is given by:

$$(6.46 - 98.6 \times 6.45)/1.35 \text{ ppm} \quad (4)$$

where 6.45 ppm is the shift of resonance B in the absence of radicals.

To illustrate that the exchange is indeed in the fast-exchange limit, we varied the exchange rate constant k_{ex} from 0 to 10⁶ M⁻¹ s⁻¹. Coalescence of resonances B and B' occurs at approximately 10⁴ M⁻¹ s⁻¹ and a superposition of the spectra obtained with $k_{ex} = 0.5 \times 10^5$ to 10⁶ M⁻¹ s⁻¹ show only negligible changes in linewidth, consistent with our suggestion that exchange is indeed in the fast regime for these k_{ex} values. k_{ex} was then set to either 0.50 × 10⁵, 1.05 × 10⁵ or 3.0 × 10⁵ M⁻¹ s⁻¹ in the simulations, which corresponds to the measured values at 283.5, 293.0 and 313.0 K from the variable-temperature experiment (Fig. 3c; extracted from the analysis of resonances A and C). As the temperature increases, the FWHM of resonance B/B' decreases from 16.7 to 13.9 Hz and then to 10.7 Hz in the simulations. The experimental spectra (Extended Data Fig. 6b) are more complex because they contain/coupling. However, deconvolution of the resonances indicates that the broadening decreases from 3.3 to 3.2 Hz and then to 2.9 Hz. On increasing the radical concentration to 9 mM, the simulated line broadenings decrease from 29 to 19 Hz with the same exchange rates, whereas experimentally the linewidth decreases from 24.1 to 23.5 Hz. Although our simulations are in reasonable agreement with experiment, they predict slightly larger broadenings than seen experimentally, and larger errors are observed with lower concentrations of radicals. This is ascribed to: (1) uncertainty in the T_1 of resonance B'—decreasing this value from 500 Hz (the value chosen for the original simulations) to 50 Hz decreases the line broadening of resonance B/B' from 13.9 to 7.7 Hz for $k_{ex} = 1.0 \times 10^5$ M⁻¹ s⁻¹ and 5 mM radical concentration; and (2) uncertainty in the shift of B as charging progresses and the pH changes—only very small changes will have a substantial effect on the line broadening, because the shift difference between B and B' is so small. To address this at least in part, we estimated the shift of B' at the beginning of charge, where pH and bulk magnetic susceptibility effects are probably smaller; and (3) the lack of inclusion of the effect of temperature on the shift of B' (and B) and their relaxation times.

Identification of the DHAQ decomposition products

The following NMR experiments were performed on an aliquot solution of 0.2 cm³ taken from the H-cell experiments at 470 h (which was the end of a charge–rest experiment), and sealed in a 5 mm NMR tube (see Methods section 'H-cell experiments in an argon glovebox').

¹H diffusion-ordered spectroscopy

Diffusion-ordered spectroscopy (DOSY) spectra were obtained with a 2D sequence for diffusion measurements using a stimulated echo and a longitudinal eddy-current decay using bipolar gradient pulses²⁹. The diffusion coefficient, D , was calculated with:

$$f(g) = I_0 e^{-\gamma g^2 \delta^2 (\Delta - \delta/3) D} \quad (5)$$

where γ is the gyromagnetic ratio of the proton (26,752 rad s⁻¹ G⁻¹) and g is the variable gradient strength ranging from 0 to 2,100 G cm⁻¹ with 16 increments of gradient strength. The length of the gradient δ is 1.5 ms and the diffusion time Δ is 100 ms.

¹H homonuclear correlation spectroscopy

Homonuclear correlation spectroscopy (COSY) with an artefact-free PFG-enhanced double quantum filter³⁰ was performed. The second dimension was constructed using 192 increments spanning 13 ppm. d_1 was 2 s and n_s for each increment was 2.

¹³C distortionless enhancement polarization transfer

Spectra were acquired with a shaped pulse of 180° on ¹³C, and a 135° read pulse and proton decoupling were applied during acquisition. The delay between the 90° and 180° pulses was 3.45 ms. This pulse

Article

sequence yields positive peaks with one or three protons bonded to the carbon atom and negative peaks with two protons bonded to the carbon atom.

¹H/¹³C correlation heteronuclear single quantum coherence

Experiments were performed via an insensitive nuclei enhanced by polarization transfer (INEPT) experiment^{31–33}. Phase-sensitive acquisition with an echo/antiecho-TPPI (time-proportional phase incrementation) gradient selection and decoupling was used during acquisition. Trim pulses for the INEPT method with multiplicity editing during the selection step as well as shaped pulses for inversion on ¹³C for matched sweep adiabatic pulses were also used during acquisition. The second dimension (¹³C) was constructed using 1,024 increments spanning 190 ppm. d_1 was 0.8 s and n_s for each increment was 4.

H-cell experiments in an argon glovebox

Study of electrolyte decomposition and battery self-discharge. The solvent was prepared by dissolving 1 M KOH in D₂O followed by vigorous degassing for 2 h, and was then loaded into a glovebox (oxygen level <0.1 ppm), along with parts of the H cell. The oxygen level of the solvent was measured to be 0 ppm (0.2 ppm accuracy) by a precision dissolved-oxygen meter. 100 mM DHAQ and 300 mM K₄Fe(CN)₆ solutions were prepared in 20 cm³ of solvent and were placed in the H cell inside the glovebox. The H cell with a sampling port was made in-house (Supplementary Fig. 3a). Pretreated Nafion membrane was sandwiched between the two half-cells. Three pieces of carbon paper (Sigracet 39AA) with a dimension of 1 cm × 3 cm were immersed in the solution on each side and electrically connected to a portable potentiostat (SP-150, BioLogic SAS). A charge current of 10 mA was applied to a cut-off voltage of 1.9 V and the voltage was continuously measured during cell resting for up to 470 h. The solution was stirred vigorously during the experiment. NMR, infrared and mass spectrometry analysis were performed on the solution and headspace gas during the charge–rest cycling.

Study of battery self-discharge by infrared spectroscopy

Attenuated total reflection infrared spectra were acquired with a Fourier-transform infrared spectrometer (Cary 630) inside an argon glovebox. The spectral range was from 4,000 to 400 cm^{−1} with a resolution of 2 cm^{−1}. Sixteen background and sample scans were performed.

During the charge–rest experiment, 10 mm³ aliquots of DHAQ solution were extracted intermittently from the sampling port of the H cell by a syringe. These were then dropped onto the spectrometer sampling window to enable infrared spectra to be acquired.

Ex situ mass spectrometry

Mass spectrometry was performed using an in-house system connected to a quadrupole gas analysis system (Pfeiffer ThermoStar). The analysis was performed on 2 cm³ of gas sampled from the headspace of the same solution in the H cell after the charge–rest cycling. 5 cm³ syringes were used to extract the gas, and the syringes containing the sampled gas were transported in an airtight plastic box from the glovebox to the mass spectrometer. The transport time was approximately 2 min. The gas or solution was injected into an online T-shape glass sampler. The carrier gas was argon at a flow rate of 100 μL s^{−1} (1 L_n, 1 litre under normal conditions of 101.3 kPa and 0 °C) at 1.1 bar(a) (absolute pressure). The dwell time for $m/z = 4$ was 5 s.

Study of gas evolution by in situ mass spectrometry

An electrochemical H cell was designed and connected to an online electrochemical mass spectrometry system (Supplementary Fig. 4a, b). The cell was based on two 1/2 inch stainless steel tees (Swagelok) interconnected via a liquid-tight glass union with a membrane fitted in the middle (Nafion 212). Both sides of the cell were capped at the bottom with a round-bottomed glass test tube fitted with a magnetic stirrer

and at the top with a stainless steel plunger where a working electrode was affixed. All connections were made both liquid- and gas-tight with PTFE ferrules. The total internal volume of each compartment was about 10 cm³. To sample the headspace (about 1 cm³) the top plunger of the anolyte (AQ) compartment of the cell was fitted with two stainless steel tubes that were connected to a gas line through double-shut-off quick connects (Beswick). The flow of the argon carrier gas was controlled by a mass-flow controller and a pressure controller (Bronkhorst) and set to 200 μL s^{−1} at 1.1 bar(a). After passing through the head space of the anolyte, the sample gas was fed to a quadrupole mass spectrometer (Pfeiffer ThermoStar) through a capillary (inner diameter = 0.22 mm) heated to 120 °C to prevent condensation. A potentiostat (Ivium Vertex) was connected to both sides of the cell to control the electrochemical operations.

Cyclic voltammetry

Electrochemical measurements were performed on a potentiostat (SP-150, BioLogic SAS) using an in-house small-volume CV cell. Polished 3 mm diameter glassy carbon (A-012744, BioLogic SAS) was used as the working electrode, and coiled platinum wire was used as the counter-electrode. For the reference, a Hg/HgO (1 M KOH) electrode with a potential of 0.14 V versus SHE was used. The 1 M KOH solution was made under inert atmosphere by the addition of degassed Millipore water to a known quantity of potassium hydroxide. DHAQ (16 mg) or DBEAQ (41.2 mg) were dissolved in the KOH solution (using 13.32 cm³ and 10 cm³ of 1 M KOH, respectively) under inert atmosphere. A sample of the solution was then extracted and added to the nitrogen-flushed small-volume electrochemical cell under inert atmosphere. The electrodes were then checked to ensure that no bubbles had formed during the addition of water before cycling was initiated. A constant overpressure of nitrogen was maintained during the experiment. The voltage was scanned from 0 V to −1.5 V at 20 mV s^{−1}.

Equilibrium concentrations of DHAQ^{2−}, DHAQ^{3−} and DHAQ^{4−}, and CV fittings

In Extended Data Fig. 4h, we have fitted the first full CV cycle with four different approaches. The two 2e[−] processes (approaches 1 and 2) are an illustrative comparison, whereas the 3rd and 4th approaches (both 1e[−] + 1e[−] processes) are more relevant. In each case, we implemented code using the SciPy Python library and its curve-fitting function. As the values that were to be fitted have differing orders of magnitude, the $x_scale = jac$ option was used to rescale variables to aid in the fits. The diffusion coefficients were assumed to be the same for each species, and where relevant so was the electron-transfer rate, k_0 . The symmetry factor α was set to 0.5. For approach 4, diffusion coefficients and electron-transfer rates were not constrained to be the same for all species. Initial guesses for the fits were set to be in line with the reported experimental data.

For approach 3, the fitted voltage difference was averaged over the 68 cycles analysed here and found to be 60 mV. This is in line with previously reported data⁵. However, these values are twice what we found from our comproportionation calculations coupled with the Evans method. The average diffusion coefficient is 2.7×10^{-10} m² s^{−1}. The average potentials (versus Hg/HgO (1 M KOH)) were −0.804 V and −0.864 V. The k_0 values had a far larger distribution in our fits, as the average of 3×10^{-3} cm s^{−1} had a standard deviation between 2×10^{-3} and 3×10^{-3} cm s^{−1}.

Approach 4 gave a fitted voltage difference of 30 mV, which matches very well with the value of 33 mV derived from the Evans method. The other fitted parameters are as follows: diffusion coefficients were 3.0×10^{-10} , 1.0×10^{-10} and 7.8×10^{-10} m² s^{−1} for DHAQ^{2−}, DHAQ^{3−} and DHAQ^{4−}, respectively. The potentials were $E_1 = -0.800$ V and $E_2 = -0.830$ V. $k_{0,1}$ and $k_{0,2}$ (the electron-transfer rate k_0 for reactions 1 and 2, respectively) converged to the same value, 7×10^{-3} cm s^{−1}.

These approaches (Extended Data Fig. 4h) show that it is far more probable that the redox reactions are two single-electron steps instead

of a single two-electron step. It is non-trivial to obtain the voltage difference $E_1 - E_2$, and inconsistent results are returned owing to the complexity of the fitting procedure. Such nonlinear least-squares fitting can be very sensitive to initial conditions and constraints used. Comparing these approaches with the comparative simplicity of using the equilibrium constant along with the Evans method shows that for $E_1 - E_2$ all the fits are reasonable. We believe that the Evans method provides an independent method for determining $E_1 - E_2$ that is—at least in the DHAQ case—associated with smaller errors (that is, it is more accurate) than the values provided purely through fitting of the CV data.

The current model does not capture the asymmetry in the HOD shift as a function of SOC, which leads to an asymmetry in the calculated DBEAQ^{3•-} radical concentrations as a function of SOC (Extended Data Fig. 4d). This suggests that either a competing reaction is present that depletes the radical concentration at higher concentrations of DBEAQ⁴⁺ or that possibly, changes in the solvation of the ions with SOC also lead to changes in the HOD chemical shift.

EPR experiment

The X-band EPR experiment was performed on a solution of 100 mM and 1 mM DHAQ, both reduced to approximately 50% of their SOC, using an EPR spectrometer (Bruker EMX). For the 100 mM DHAQ, the field was swept from 3,000 to 4,000 G and the microwave frequency was 9.865410 GHz. For the 1 mM DHAQ sample, the field was swept from 3,463 to 3,563 G and the microwave frequency was 9.865408 GHz. For both experiments, the attenuation was 20 dB, the amplitude of modulation was 0.1 G, the power was 2 mW, and the sweep time was 20 s. A sampling tube (30 mm³ cm⁻¹, Bruker AquaX) was used for data acquisition.

The software package EasySpin was used to fit the spectra³⁴. Spin systems of one unpaired electron coupled to different numbers of proton and deuterium spins were set up, taking into account the H–D exchange in deuterated solvent D₂O. For the non-deuterated DHAQ^{3•-}, a spin system of three protons with two magnetically equivalent spins for each proton was modelled. For singly deuterated DHAQ^{3•-}, a spin system of two protons was set up that had with two magnetically equivalent spins for each proton, another proton with a single spin and a deuteron with a single spin. For the doubly deuterated DHAQ^{3•-}, a spin system of two protons was set up that had with two magnetically equivalent spins for each proton and one deuteron with two magnetically equivalent spins. The fitted variables were g-factors, hyperfine coupling constants, the fractions of each component and linewidth. The Nelder–Mead simplex method was used. In Extended Data Fig. 3b–d, we present the fitted EPR spectra using a single component of non-deuterated DHAQ^{3•-}, two components of non-deuterated and singly deuterated DHAQ^{3•-}, and two components of non-deuterated and doubly deuterated DHAQ^{3•-}. The two-component system comprising non-deuterated (92.6%) and doubly deuterated (7.4%) DHAQ^{3•-} gives the lowest root-mean-square deviation, 0.0175. The fit yields a g-factor of 2.0077 and hyperfine coupling constants of 0.15 MHz, 2.63 MHz and 4.64 MHz for each unique proton. The other two fits yield similar results. The difference between the g-factors of the 100 mM (2.0036) and 1 mM (2.0077) solutions is possibly due to motional effects from the varying viscosity of the solution or the magnetic field drift of the instrument.

To understand the effect of the hyperfine coupling on the NMR chemical shift, we need to correlate the paramagnetic component of the shift δ to the isotropic (Fermi-contact) hyperfine coupling constant, A_{iso} . By definition, A_{iso} describes the strength of the electron–nuclear spin coupling (that is, the unpaired spin density at the nucleus) in the limit of static (non-flipping) electronic spins. However, at finite temperatures the paramagnetic behaviour of the electrons means that they undergo a rapid flipping between the two spin states of a spin-1/2 system. In an EPR experiment where the electronic spin transitions are observed on the timescale of picoseconds to nanoseconds, this results in a coupling which is evidenced (at least in the dilute system)

by splitting of the resonance by A_{iso} . However, in NMR experiments, the much longer timeframe of the nuclear spin transitions (microseconds) results in a decoupling of the electronic transitions from the nuclear transitions. The net effect of this is that only the time-averaged electronic spin moment is felt by the nuclear spin, and the strength of this coupling determines the observed Fermi-contact NMR shift. This scaling is typically done by means of the magnetic susceptibility χ , which, for a Curie–Weiss system, depends inversely on the temperature T . The isotropic component of the total shift, δ_{iso} , can then be written as a sum of the diamagnetic (chemical shift) and paramagnetic (Fermi-contact) components: $\delta_{\text{iso}} = \delta_{\text{CS}} + \delta_{\text{FC}}$. The temperature dependence of δ_{FC} is expressed as (in ppm)

$$\delta_{\text{FC}} = A_{\text{iso}} \left(\frac{\gamma_e}{\gamma_H} \right) \left(\frac{h}{4k_B T} \right) \times 10^6 \quad (6)$$

where γ_e (γ_H) is the electron (proton) gyromagnetic ratio, h is the Planck constant, and k_B is the Boltzmann constant. The expected shifts determined from the DFT calculations and via EPR are shown in Extended Data Fig. 3e, along with A_{iso} for the DHAQ^{3•-} radical and the corresponding EPR data.

Although the EPR investigation of this redox system is still ongoing and subject to a separate study, the preliminary results clearly show why the ¹H NMR shift of proton B on reduction is relatively small. The disappearance of the ¹H NMR resonances of protons A and C on reduction is ascribed to the much larger hyperfine shifts associated with these protons (see Fig. 3d). We have not accounted for any pseudocontact contributions to the NMR shifts, because the effective g-factor is very close to the free-electron value, suggesting that they are extremely small.

Application considerations in redox flow batteries and beyond

Organic redox flow batteries based on inexpensive (for example, the estimated price of anthraquinone disulfonic acid is currently in the range of US\$0.90 per kg to US\$3.90 per kg for industrial-scale production)³⁵ and sustainable redox-active materials are promising storage technologies, which are cheaper and have fewer environmental hazards compared to the more established and mature vanadium-based systems; the price of vanadium pentoxide in Europe in 2018 ranged between US\$19.40 per kilogram to US\$63.50 per kilogram (<http://www.vanadiumprice.com>).

Because of the simplicity of the on-line NMR setup, which in essence consists of a laboratory-scale redox flow battery and a flow NMR sampling tube, we expect adoption of this technique to advance the understanding of various redox chemistries, such as quinone-based^{36,37}, carbonyl-nitrogen-based³⁸, radical-based³⁹, polymer-based⁴⁰, and metal-complex-based⁴¹ redox chemistries in flow and other battery systems^{3–5,38,42–51}—for example, lithium–air batteries that involve organic redox shuttles. This technique can be readily coupled with other (flow) characterizations, including in situ mass spectrometry, EPR and optical methods. In addition to the study of redox chemistry, the on-line technique can be exploited to study the rate of electrolyte crossover, which would help improve membrane design. The operando design lays the foundation for future magnetic resonance imaging experiments to monitor flow, and the electrolyte/solvent distribution in the electrode.

The Evans method enables the radical concentration to be determined from the magnetic susceptibility. This affords a straightforward approach to track the SOC of the anolyte and catholyte, providing critical information about cell balancing and how it varies with cycle life. This is not easy information to determine from full-cell measurements without the use of a reference electrode. Of note, our results also motivate the development of simpler and cheaper methods to extract this information by using a magnetometer or a relaxometer. We anticipate that the in situ NMR and related metrologies will contribute to our fundamental and practical understanding of flow batteries

and thus to the development of longer-lasting devices with higher energy densities for large-scale energy storage.

Data availability

The data that support the findings of this study are available from <https://doi.org/10.17863/CAM.46274> and the corresponding author upon reasonable request.

27. Dalitz, F., Cudaj, M., Maiwald, M. & Guthausen, G. Process and reaction monitoring by low-field NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **60**, 52–70 (2012).
28. Hogben, H. J., Krzystyniak, M., Charnock, G. T. P., Hore, P. J. & Kuprov, I. Spinach – a software library for simulation of spin dynamics in large spin systems. *J. Magn. Reson.* **208**, 179–194 (2011).
29. Johnson, C. S., Jr. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 203–256 (1999).
30. Shaw, A. A., Salaun, C., Dauphin, J.-F. & Ancian, B. Artifact-free PFG-enhanced double-quantum-filtered COSY experiments. *J. Magn. Reson. A* **120**, 110–115 (1996).
31. Willker, W., Leibfritz, D., Kerssebaum, R. & Bermel, W. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **31**, 287–292 (1993).
32. Zwahlen, C. et al. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119**, 6711–6721 (1997).
33. Boyer, R. D., Johnson, R. & Krishnamurthy, K. Compensation of refocusing inefficiency with synchronized inversion sweep (CRISIS) in multiplicity-edited HSQC. *J. Magn. Reson.* **165**, 253–259 (2003).
34. Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **178**, 42–55 (2006).
35. Dieterich, V. et al. Estimating the cost of organic battery active materials: a case study on anthraquinone disulfonic acid. *Transl. Mater. Res.* **5**, 034001 (2018).
36. Yang, Z. et al. Alkaline benzoquinone aqueous flow battery for large-scale storage of electrical energy. *Adv. Energy Mater.* **8**, 1702056 (2018).
37. Gerhardt, M. R. et al. Anthraquinone derivatives in aqueous flow batteries. *Adv. Energy Mater.* **7**, 1601488 (2017).
38. Orita, A., Verde, M. G., Sakai, M. & Meng, Y. S. A biomimetic redox flow battery based on flavin mononucleotide. *Nat. Commun.* **7**, 13230 (2016).
39. Wei, X. et al. TEMPO-based catholyte for high-energy-density nonaqueous redox flow batteries. *Adv. Mater.* **26**, 7649–7653 (2014).
40. Janoschka, T. et al. An aqueous, polymer-based redox-flow battery using non-corrosive, safe and low-cost materials. *Nature* **527**, 78–81 (2015).
41. Chen, J.-J., Symes, M. D. & Cronin, L. Highly reduced and protonated aqueous solutions of $[\text{P}_2\text{W}_{18}\text{O}_{62}]^{6-}$ for on-demand hydrogen generation and energy storage. *Nat. Chem.* **10**, 1042–1047 (2018).
42. Winsberg, J., Hagemann, T., Janoschka, T., Hager, M. D. & Schubert, U. S. Redox-flow batteries: from metals to organic redox-active materials. *Angew. Chem. Int. Ed.* **56**, 686–711 (2017).
43. Kowalski, J. A., Su, L., Milshtein, J. D. & Brushett, F. R. Recent advances in molecular engineering of redox active organic molecules for nonaqueous flow batteries. *Curr. Opin. Chem. Eng.* **13**, 45–52 (2016).
44. Kwon, G. et al. Multi-redox molecule for high-energy redox flow batteries. *Joule* **2**, 1771–1782 (2018).
45. DeBruler, C. et al. Designer two-electron storage viologen anolyte materials for neutral aqueous organic redox flow batteries. *Chem* **3**, 961–978 (2017).
46. Service, R. F. Advances in flow batteries promise cheap backup power. *Science* **362**, 508–509 (2018).
47. Ji, Y. et al. A phosphonate-functionalized quinone redox flow battery at near-neutral pH with record capacity retention rate. *Adv. Energy Mater.* **9**, 1900039 (2019).
48. Huskinson, B. et al. A metal-free organic–inorganic aqueous flow battery. *Nature* **505**, 195–198 (2014).
49. Ding, Y., Zhang, C., Zhang, L., Zhou, Y. & Yu, G. Molecular engineering of organic electroactive materials for redox flow batteries. *Chem. Soc. Rev.* **47**, 69–103 (2018).
50. Holland-Cunz, M. V., Cording, F., Friedl, J. & Stimming, U. Redox flow batteries—concepts and chemistries for cost-effective energy storage. *Front. Energy* **12**, 198–224 (2018).
51. Ding, Y., Li, Y. & Yu, G. Exploring bio-inspired quinone-based organic redox flow batteries: a combined experimental and computational study. *Chem* **1**, 790–801 (2016).
52. Ross, R. T. Dipolar broadening of EPR spectra due to solute segregation in frozen aqueous solutions. *J. Chem. Phys.* **42**, 3919–3922 (1965).

Acknowledgements E.W.Z. and C.P.G. acknowledge support from Centre of Advanced Materials for Integrated Energy Systems (CAM-IES), via EPSRC grant number EP/P007767/1. E.W.Z., R.J. and C.P.G. acknowledge support from Shell. E.W.Z. acknowledges support from the Manifest exchange programme via EPSRC grant number EP/N032888/1. T.L. acknowledges support from the Schlumberger Fellowship, Darwin College. E.J. acknowledges support from the Swedish Research Council. We thank A. Brookfield for assistance with the EPR measurement; P. A. A. Klusener from Shell, H. Bronstein, I. Fleming, D. S. Wright, K. Märker, C. Xu, P. C. M. Magusin from the University of Cambridge and E. Castillo-Martínez from Universidad Complutense de Madrid for discussions; R. Tan from Imperial College London and D. Lyu, Y. Kim, Y. Jin, and J. Lu from the University of Cambridge for assistance setting up the redox flow battery. A.W. and Q.S. acknowledge Imperial College start-up funding and CAM-IES seed funding. J.C.G. acknowledges support from the Spanish Ministry of Science, Innovation and Universities through a Ramon y Cajal Fellowship (RYC-2015-17722) and the Retos Project (MAT2017-86796-R).

Author contributions C.P.G. supervised the project. E.W.Z. and C.P.G. conceived the idea. E.W.Z. designed the in situ setups and performed the NMR and EPR experiments and analysis. T.L. and E.W.Z. performed the infrared experiments. E.J. performed the calculations of reaction equilibrium, CV fittings and NMR chemical shift. J.L. performed the DFT calculations of the radical species and their proton hyperfine couplings. I.T. and E.W.Z. designed and performed the in situ mass spectrometry with assistance from T.L. R.J. performed the CV experiments. A.W. synthesized DBEAQ under the supervision of Q.S. H.S. and J.C.G. assisted in performing the in situ NMR experiments. All authors contributed to the discussion of the project. E.W.Z. and C.P.G. wrote the manuscript with input from all co-authors.

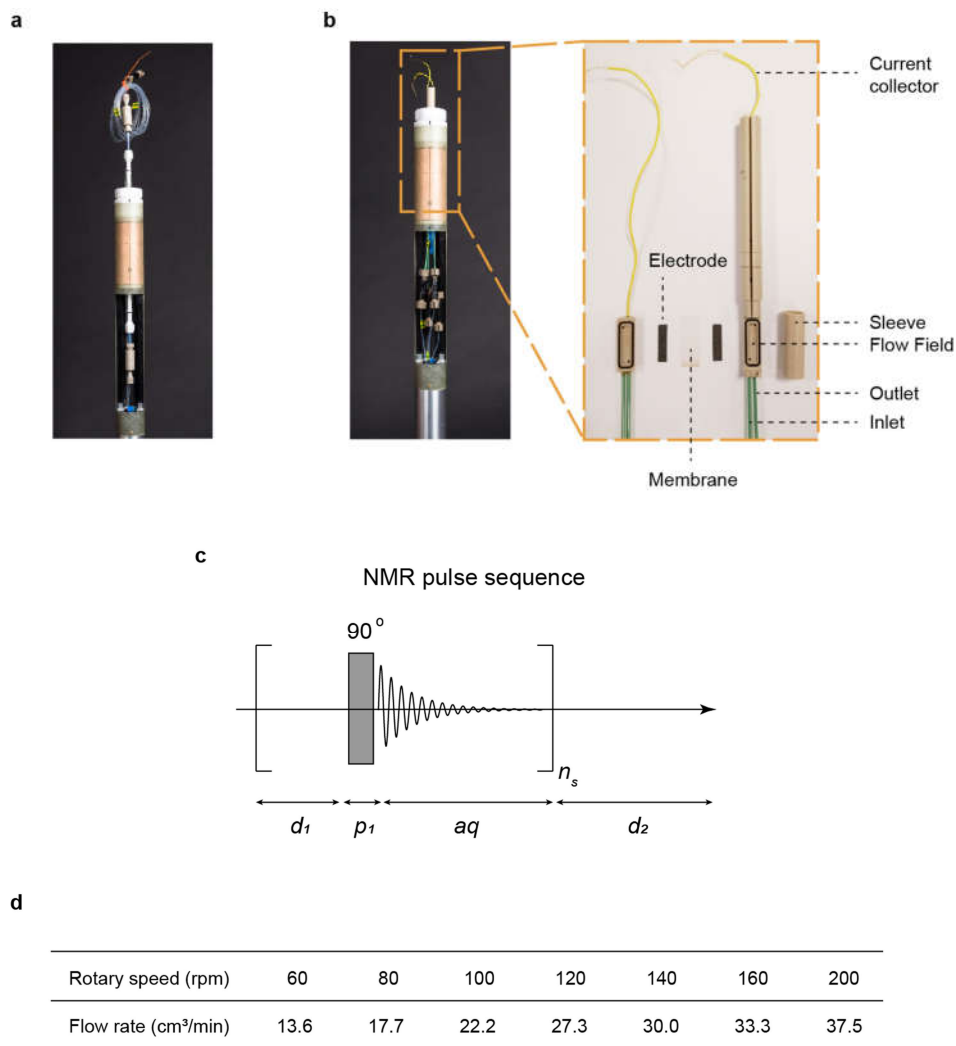
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2081-7>.

Correspondence and requests for materials should be addressed to C.P.G.

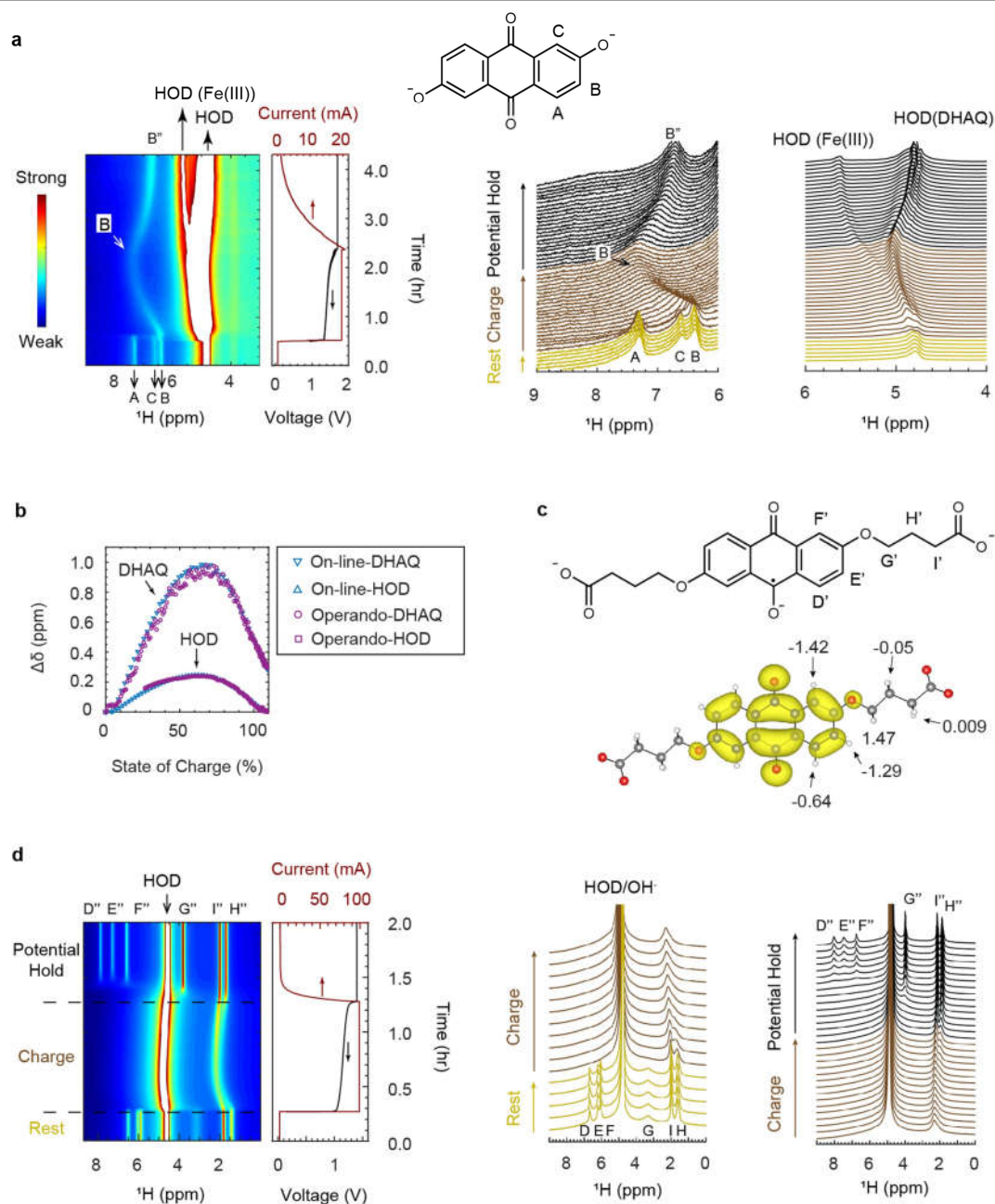
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | The in situ NMR probe setups and pulse sequence.

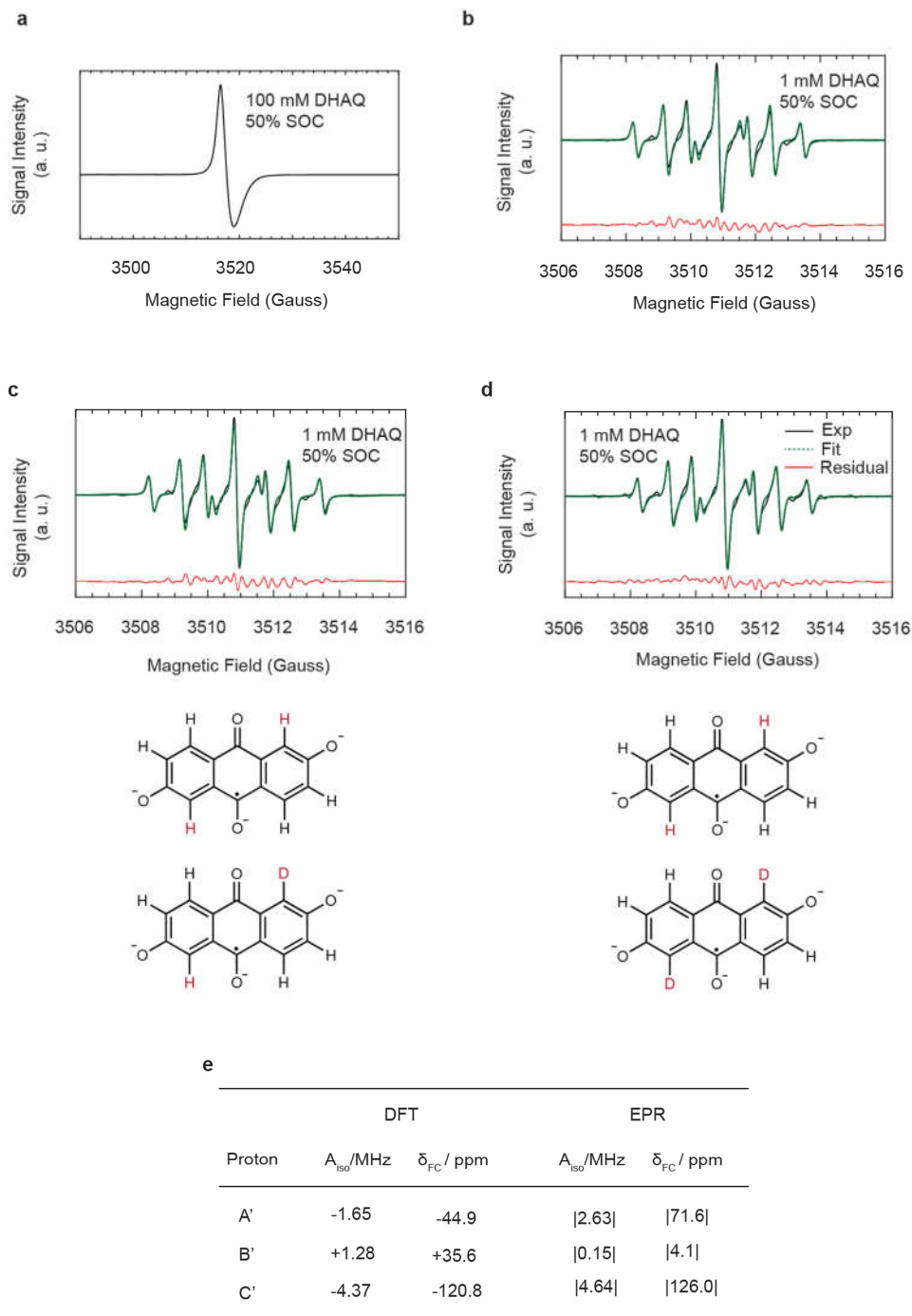
a, Image of the flow-through (on-line) sampling tube positioned inside a micro-imaging probe (Bruker 2.5). **b**, The operando cell assembly positioned inside the NMR probe. Inset, magnification of the components of the cell assembly. **c**, Each spectrum is acquired by collecting n_s free induction decays (FIDs; that

is, scans) with a recycle delay of d_1 . A second time delay, d_2 , is introduced between each spectrum. The pulse width, p_1 , for a 90° pulse is $27 \mu\text{s}$. The acquisition time for each FID (aq) is 1.5 s. **d**, Measured flow rates as a function of rotary speed of the pump.



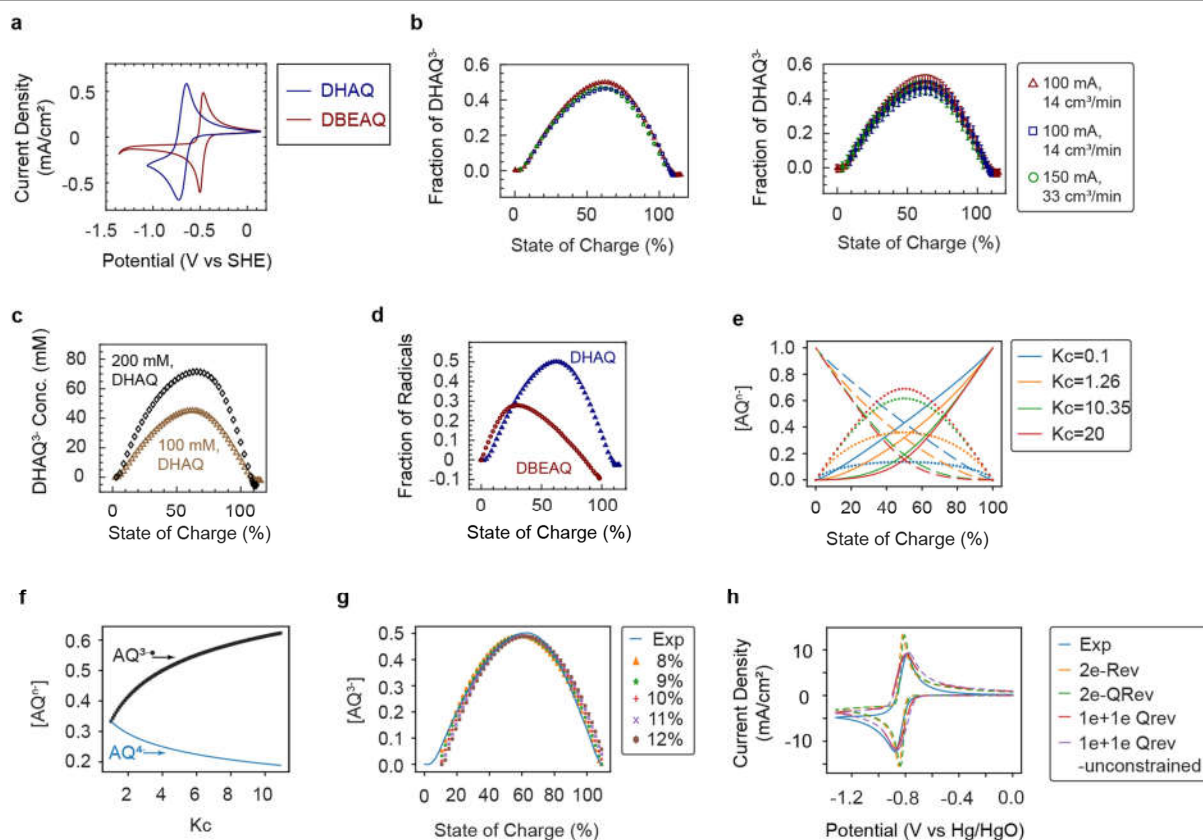
Extended Data Fig. 2 | The in situ pseudo-2D and stacked ^1H NMR spectra versus battery voltage and current. **a**, Operando NMR spectra of 10 cm^3 of 100 mM DBEAQ, against 20 cm^3 of 150 mM potassium ferrocyanide. The FWHM of the water signal acquired via on-line detection is 6 Hz , compared to the FWHM of 59 Hz acquired in the operando setup. **b**, Shift of the DHAQ- H^{B} and water resonances ($\Delta\delta$) as a function of SOC measured via on-line and operando detection. The shifts of the water resonances were only included in the plot

after there was clear peak separation between the water signals arising from the catholyte and anolyte electrolytes. **c**, Labelling of the protons, and DFT-derived volumetric plots of the singly occupied molecular orbital of DBEAQ $^{3-}$. The values of the isotropic Fermi contact hyperfine coupling constants (in MHz) were derived from the DFT calculation. **d**, On-line NMR spectra of 20 cm^3 of 100 mM DBEAQ (anolyte), against 20 cm^3 of 300 mM potassium ferrocyanide (catholyte).



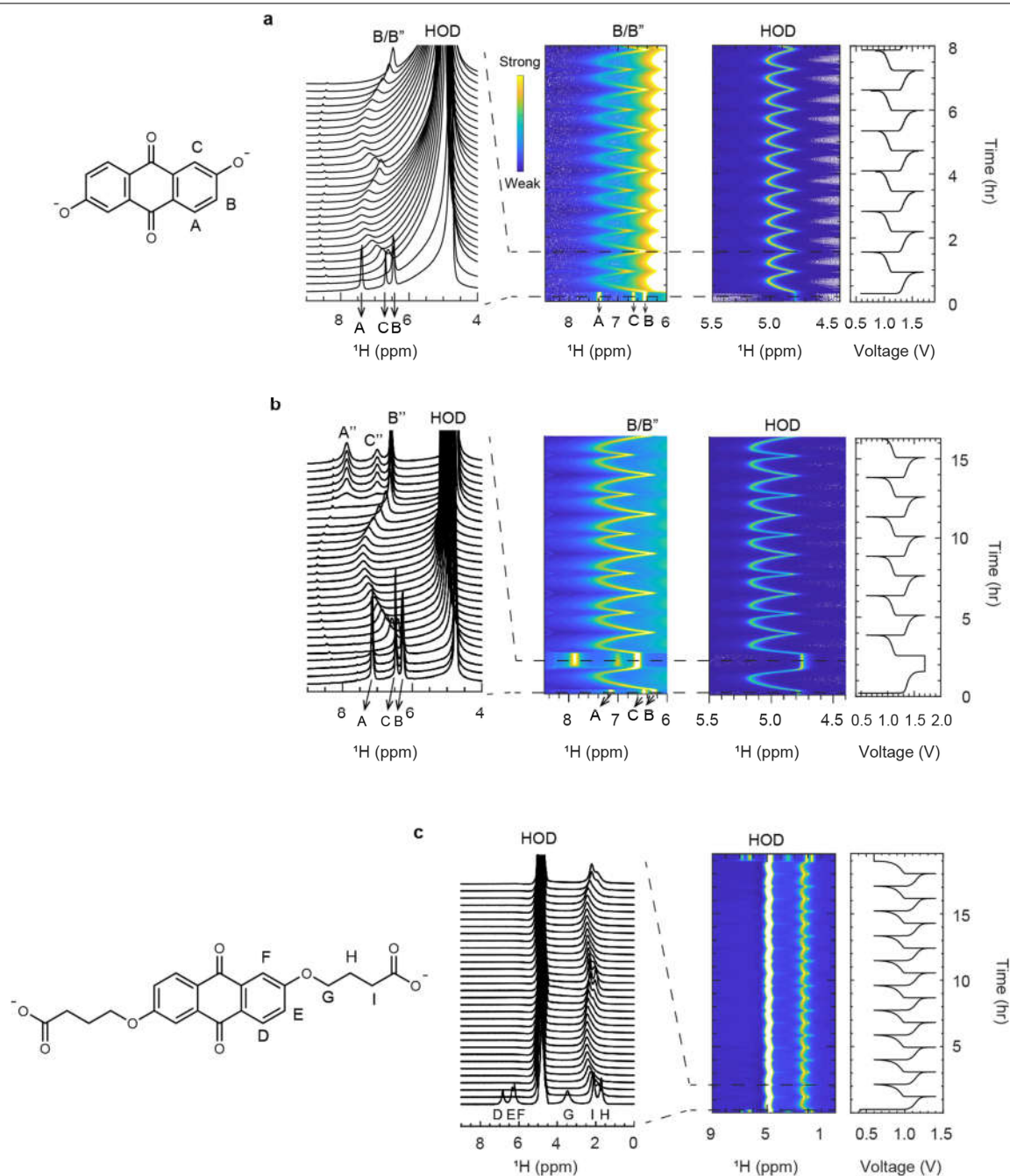
Extended Data Fig. 3 | EPR spectra of DHAQ^{3•-} radical anions. a–d, Spectra of 100 mM (a) and 1 mM (b–d) DHAQ solution reduced to 50% SOC. The 1 mM solution spectra were fitted with a single component comprising non-deuterated DHAQ^{3•-} (b), two components comprising non-deuterated and singly deuterated DHAQ^{3•-} (c), and two components comprising non-deuterated and doubly deuterated DHAQ^{3•-} (d). The labels used to describe the three proton positions are given in the structure of the radical shown in b. The residual is the difference between the experimental and fitted spectra. At 100 mM, an intense EPR signal of DHAQ^{3•-} was observed at 3,518 G (g-factor,

2.0036). However, the signal is strongly broadened by a combination of the electron dipolar interactions at such high radical concentrations and the rapid electron-transfer reactions. This leads to the loss of the hyperfine coupling features⁵². At 1 mM, hyperfine coupling features were resolved. e, Table showing DFT- and EPR-derived isotropic hyperfine coupling constants (A_{iso}) and contact shifts (δ_{FC}) of the proton resonances. The source of the difference between the DFT- and EPR-derived hyperfine coupling constants is probably due to solvent (water) coordination, H–D exchange, and/or electron hopping.



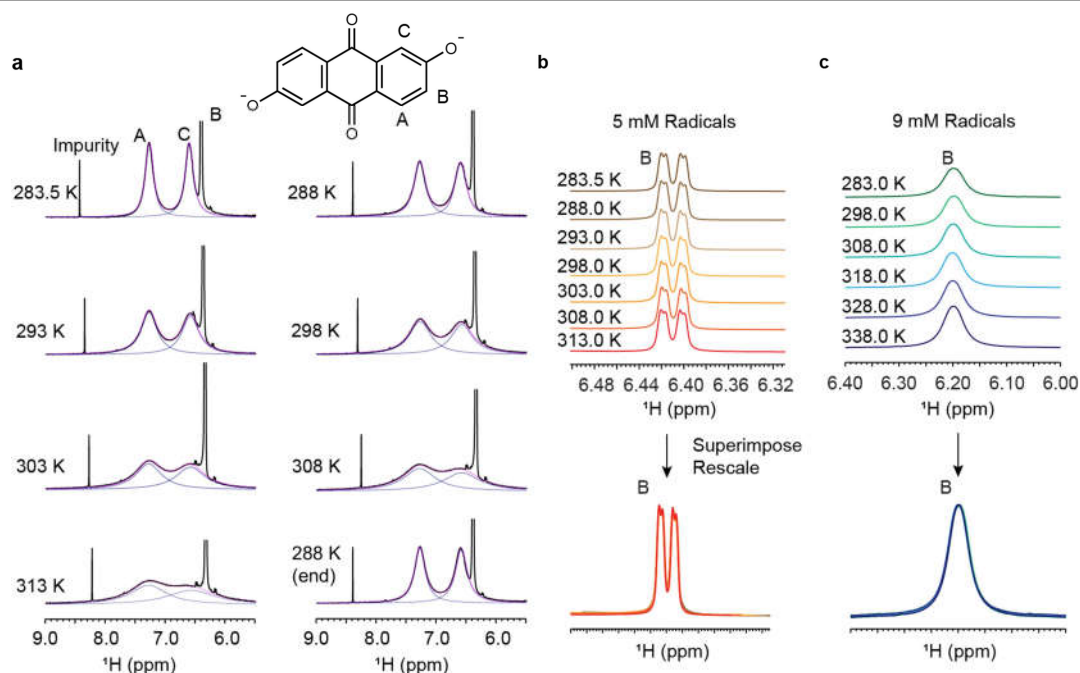
Extended Data Fig. 4 | CV and concentration of radical anions. **a**, CV of 5 mM DHAQ and 5 mM DBEAQ in D₂O with 1 M KOH. The potential was scanned at 20 mV s⁻¹, the scan starting by sweeping from positive to negative potentials. **b**, Fraction of DHAQ³⁻ radicals calculated on the basis of the Evans method as a function of SOC for DHAQ (100 mM) under a charging current of 100 mA and a flow rate of 13.6 cm³ min⁻¹ (measured twice; red and blue), and a charging current of 150 mA and a flow rate of 33.3 cm³ min⁻¹ (green). The figures on the right and left are the same data plotted without and with error bars. The errors are \pm half the FWHM of the water signal. **c**, Concentration of DHAQ³⁻ radicals as a function of SOC for 100 mM and 200 mM DHAQ. **d**, Fraction of radicals as a function of SOC for 100 mM DHAQ and 100 mM DBEAQ. **e**, Relative concentrations of AQ²⁻, AQ³⁻ and AQ⁴⁻ as a function of SOC of the system and the equilibrium constant, K_c . AQ²⁻, AQ³⁻ and AQ⁴⁻ are represented by dashed, dotted and solid lines, respectively. $K_c = 1.26$ and $K_c = 10.35$ correspond to values derived for DBEAQ and DHAQ, respectively (corresponding to $E_1 - E_2$

values of 6 mV and 60 mV)^{5,13}. Curves corresponding to $K_c = 0.1$ and $K_c = 20$ are also shown, to illustrate the effect using the smaller and larger values $E_1 - E_2 = -58.3$ mV and 76.6 mV, respectively. The former negative value illustrates the case in which the radical is strongly disfavoured. **f**, The AQ³⁻ and AQ⁴⁻ concentration at 50% SOC, as a function of K_c . AQ³⁻ and AQ⁴⁻ are the black dotted and blue solid lines, respectively. **g**, Experimentally derived radical concentrations as a function of SOC. Supplementary Information equations S6 and S7 were used to fit the experimental data, along with an SOC lag parameter ranging from 8% to 12%, where x is the fraction of AQ³⁻ for a given number of electrons n (discussed, as is the derivation of the equations, in the Supplementary Information). The fit to the data was done in SOC steps of 1%. **h**, The CV of a 100 mM DHAQ. '2e-Rev' and '2e-QRev' refer to a two electron reversible and two-electron quasi-reversible model, respectively, and the '1e+1e' curves (with and without fitting constraints) refer to a two-step, single-electron, quasi-reversible process.



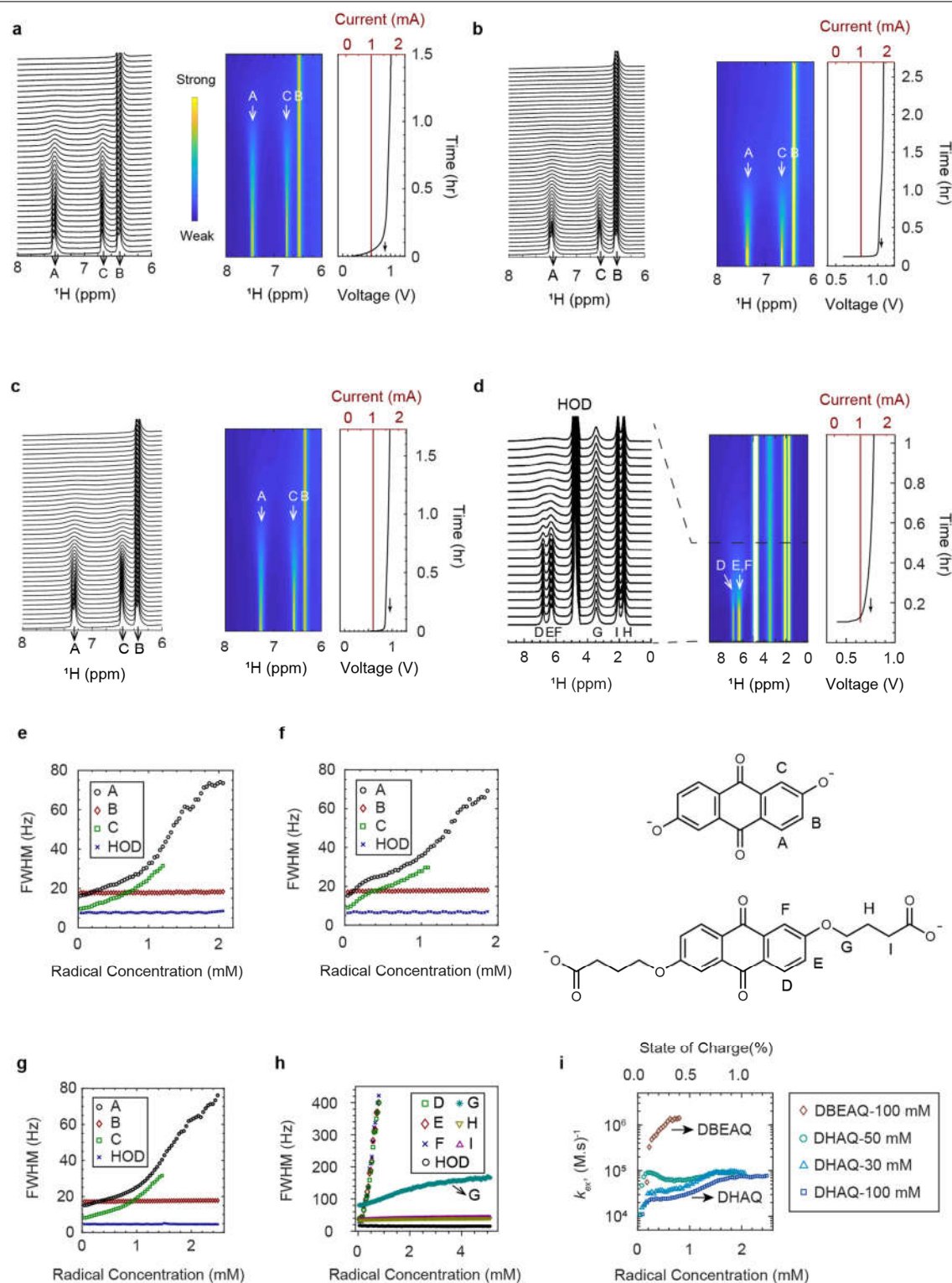
Extended Data Fig. 5 | On-line ^1H NMR spectra and battery electrochemical profile during galvanostatic cycling. **a**, 20 cm^3 of 100 mM DHAQ against 40 cm^3 of 150 mM potassium hexacyanoferrate(II)/ 50 mM potassium hexacyanoferrate(III). The flow rate is $33.3\text{ cm}^3\text{ min}^{-1}$ and the current is 150 mA . With $d_1 = 20\text{ s}$, an acquisition time (aq) = 1.5 s and $n_s = 4$, each spectrum is a snapshot of the electrochemical processes averaged over 2.5% of the SOC. **b**, 20 cm^3 of 200 mM DHAQ against 40 cm^3 of 250 mM potassium hexacyanoferrate(II)/ 50 mM potassium hexacyanoferrate(III). The flow rate is

$13.6\text{ cm}^3\text{ min}^{-1}$ and the current is 150 mA . With $d_1 = 15\text{ s}$, $aq = 1.5\text{ s}$ and $n_s = 8$, each spectrum is a snapshot of the electrochemical processes averaged over 2.9% of the SOC. **c**, 20 cm^3 of 100 mM MDBEAQ against 40 cm^3 of 150 mM potassium hexacyanoferrate(II)/ 50 mM potassium hexacyanoferrate(III). The flow rate is $13.6\text{ cm}^3\text{ min}^{-1}$ and the current is 100 mA . With $d_1 = 7\text{ s}$, $aq = 1.5\text{ s}$ and $n_s = 8$, each spectrum is a snapshot of the electrochemical processes averaged over 2.1% of the SOC.



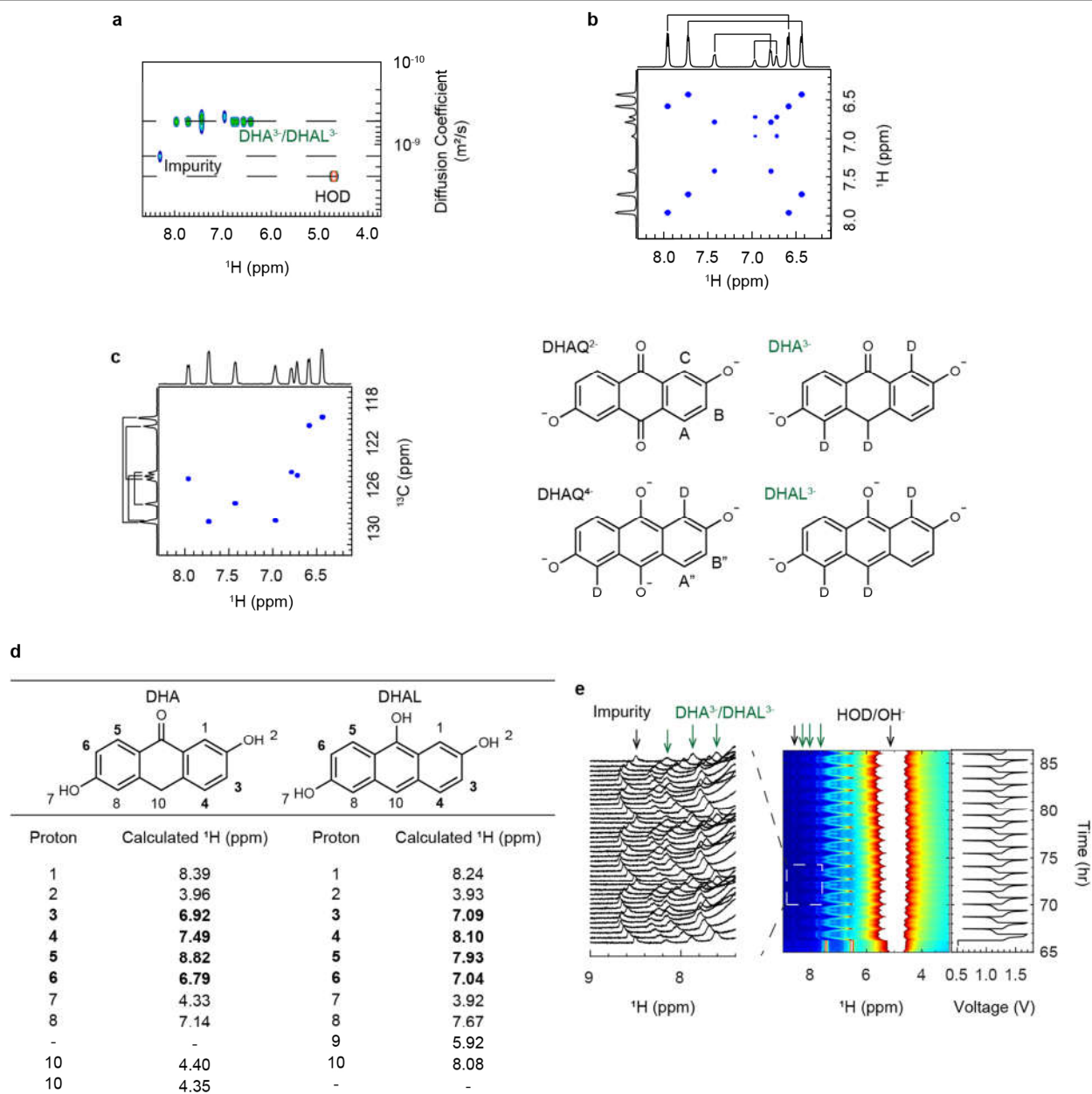
Extended Data Fig. 6 | Variable-temperature NMR of DHAQ²⁻/DHAQ³⁻ radical anion solution. **a**, Variable-temperature NMR spectra acquired from a solution containing a radical concentration of 5 mM in 100 mM DHAQ. Black, experimental spectra; navy, deconvoluted spectra; magenta: fitted-sum spectra. **b, c**, Variable-temperature NMR spectra of peak B with 5 mM (**b**), and

9 mM (**c**) of radicals. As the experiments were performed by locking the magnetic field on the shift of the water resonance (which is temperature-dependent), the shift of peak A was manually set to 7.3 ppm so as to more readily compare the different spectra.



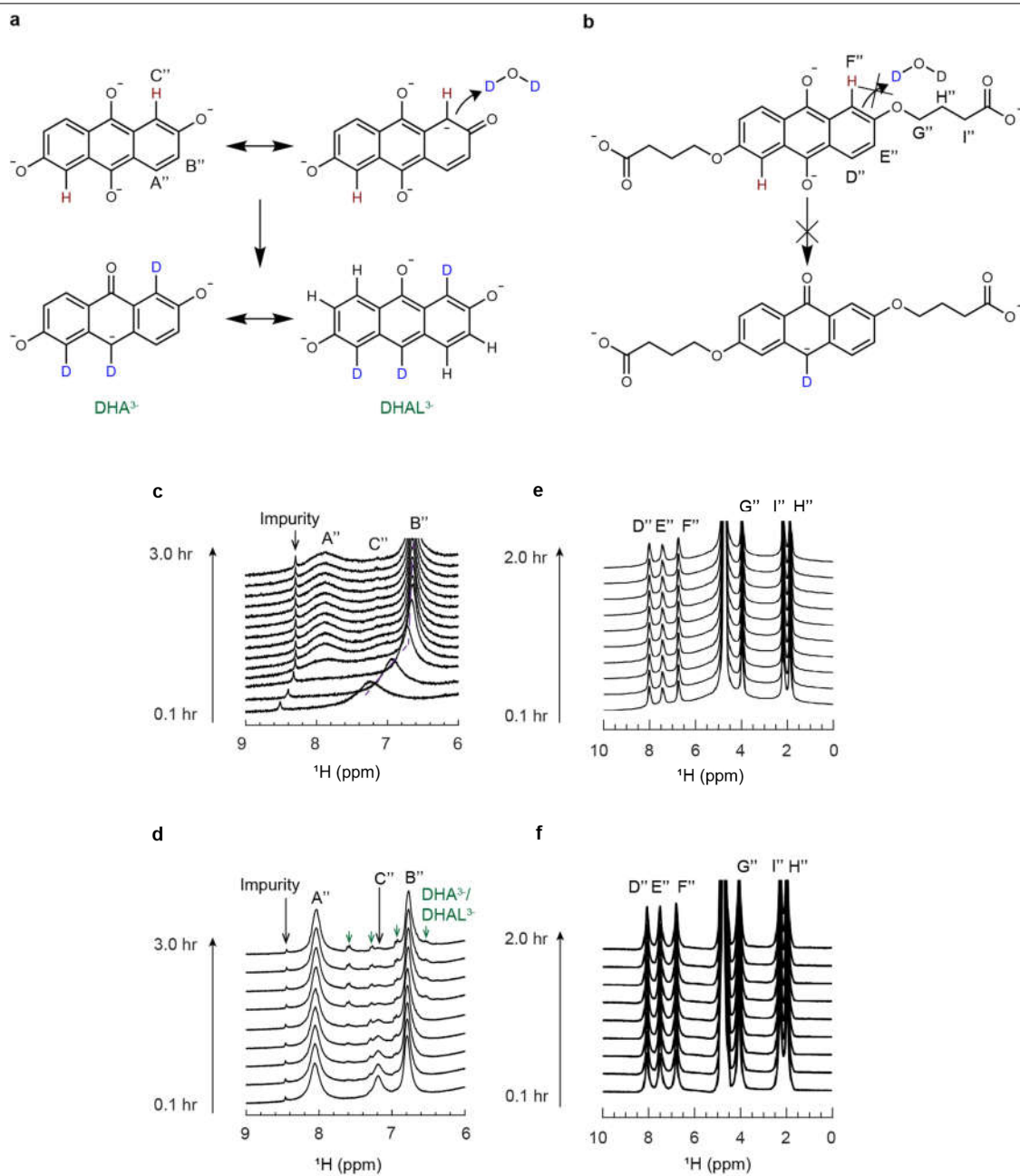
Extended Data Fig. 7 | On-line ^1H NMR spectra performed to quantify the electron-transfer rates. a–d, 30 mM (a), 50 mM (b), 100 mM 2,6-DHAQ (c), and 100 mM 2,6-DBEAQ (d) during charge/reduction at 1 mA. e–h, The FWHM of the proton signals A, B and C of DHAQ (e–g) and D, E, F, G, H and I of DBEAQ (h) as a

function of the radical concentration. i, The electron-transfer rate constant, k_{et} , calculated for 30 mM, 50 mM and 100 mM DHAQ, and 100 mM DBEAQ, at different radical concentrations and SOC, under the slow-exchange approximation.



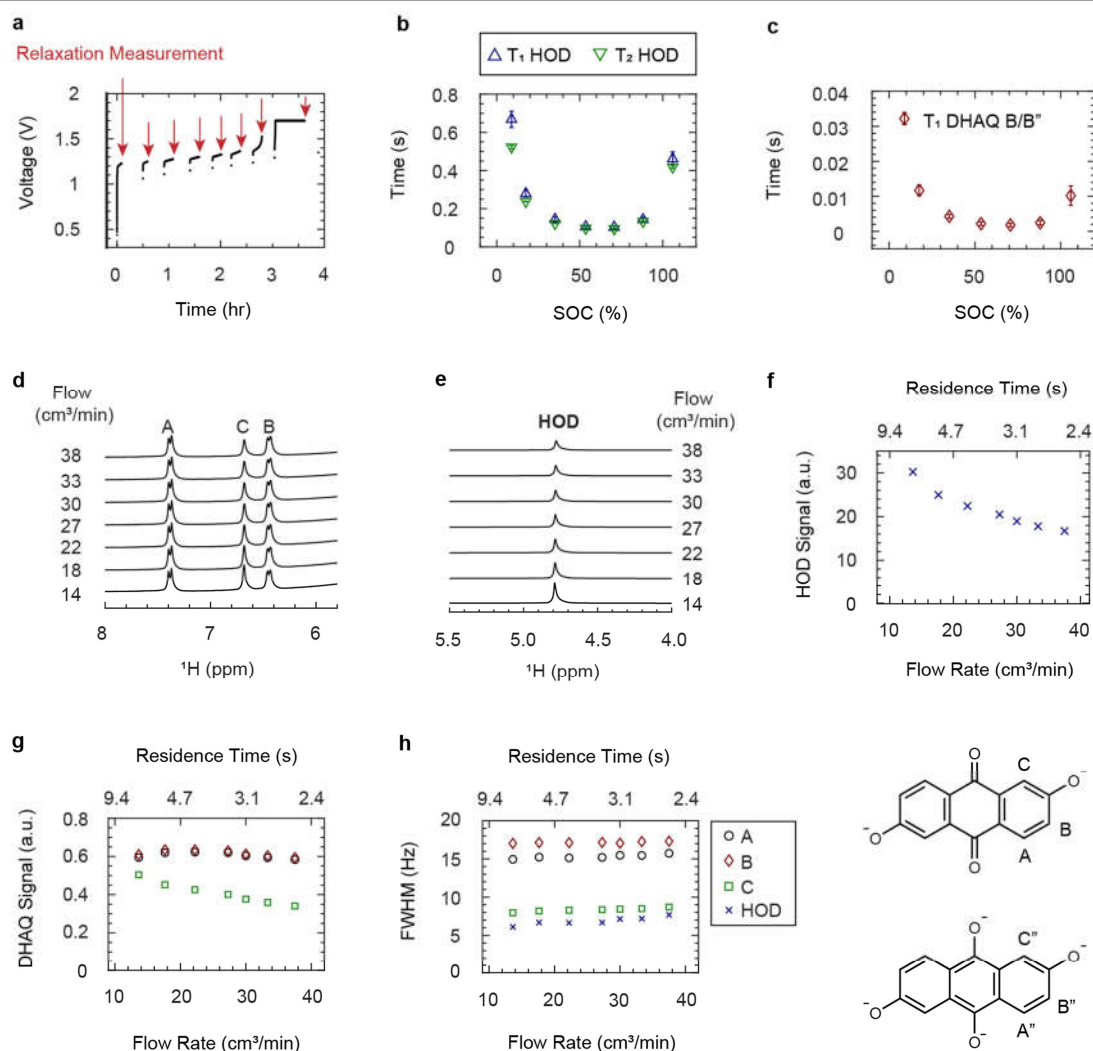
Extended Data Fig. 8 | NMR and DFT calculations to identify the decomposition products. a–c, DOSY (**a**), COSY (**b**) and heteronuclear single quantum coherence (HSQC) (**c**) spectra of the aliquot extracted from the H cell at 470 h (see Methods section ‘H-cell experiments in an argon glovebox’ for details) in an airtight NMR tube. The lines above the COSY and HSQC spectra

indicate the connectivity of the H and C atoms. **d**, DFT-derived (calculated) proton chemical shift of DHA and DHAL. **e**, In situ NMR during the galvanostatic cycling at 100 mA after the potential-hold (1.7 V) step. Green arrows, DHA^{3-} or DHAL^{3-} . The signal at 8.4 ppm arises from an impurity.



Extended Data Fig. 9 | In situ ^1H NMR study of the decomposition reaction under different cycling conditions. a, b, Proposed H-D exchange (top) and decomposition (bottom) reaction and products of DHAQ^{4-} (a) and DBEAQ^{4-} (b).

c, d, NMR spectra of DHAQ during potential hold at 1.4 V (c) and 1.7 V (d). Green arrows, DHA^{3-} or DHAL^{3-} . e, f, NMR spectra of DBEAQ during potential hold at 1.4 V (e) and 1.7 V (f). The signal at 8.4 ppm arises from an impurity.



Extended Data Fig. 10 | Effect of radicals and flow on longitudinal and transverse relaxation. **a**, The voltage profile during intermittent charging at 100 mA of 20 cm³ of 100 mM DHAQ against 40 cm³ of 200 mM K₄Fe(CN)₆. The red arrows indicate the times when the flow and electrochemical cycling were paused, and the relaxation measurements commenced. After the measurements finished, the flow and electrochemical cycling were resumed. **b**, T_1 (blue) and T_2 (green) relaxation times of the HOD signal as a function of SOC.

c, T_1 relaxation time of the DHAQ proton B/B'' signal as a function of SOC. The error bars represent the 95% confidence level from the fit. **d**, **e**, Proton NMR spectra of DHAQ (H^A, H^B and H^C; **d**) and HOD (**e**) acquired at different flow rates via on-line detection. **f**, **g**, Signal integral of HOD (**f**) and DHAQ (**g**) as a function of flow rate and residence time in the detection region of the NMR probe. **h**, The FWHM of the HOD and DHAQ signals as a function of flow rate and residence time, and the labelling of protons.

Limits on gas impermeability of graphene

<https://doi.org/10.1038/s41586-020-2070-x>

Received: 16 June 2019

Accepted: 19 December 2019

Published online: 11 March 2020

 Check for updates

P. Z. Sun^{1,2}, Q. Yang^{1,2}, W. J. Kuang¹, Y. V. Stebunov^{1,2}, W. Q. Xiong³, J. Yu⁴, R. R. Nair², M. I. Katsnelson⁴, S. J. Yuan^{3,4}, I. V. Grigorieva¹, M. Lozada-Hidalgo¹, F. C. Wang^{1,2,5} & A. K. Geim^{1,2}✉

Despite being only one-atom thick, defect-free graphene is considered to be completely impermeable to all gases and liquids^{1–10}. This conclusion is based on theory^{3–8} and supported by experiments^{1,9,10} that could not detect gas permeation through micrometre-size membranes within a detection limit of 10^5 to 10^6 atoms per second. Here, using small monocrystalline containers tightly sealed with graphene, we show that defect-free graphene is impermeable with an accuracy of eight to nine orders of magnitude higher than in the previous experiments. We are capable of discerning (but did not observe) permeation of just a few helium atoms per hour, and this detection limit is also valid for all other gases tested (neon, nitrogen, oxygen, argon, krypton and xenon), except for hydrogen. Hydrogen shows noticeable permeation, even though its molecule is larger than helium and should experience a higher energy barrier. This puzzling observation is attributed to a two-stage process that involves dissociation of molecular hydrogen at catalytically active graphene ripples, followed by adsorbed atoms flipping to the other side of the graphene sheet with a relatively low activation energy of about 1.0 electronvolt, a value close to that previously reported for proton transport^{11,12}. Our work provides a key reference for the impermeability of two-dimensional materials and is important from a fundamental perspective and for their potential applications.

From a theoretical standpoint, monolayer graphene poses a very high energy barrier for the penetration of atoms and molecules. Density functional theory (DFT) calculations predict that the energy barrier E is at least several electronvolts^{2–6}, which should prohibit any gas permeation under ambient conditions. Indeed, one can estimate that at room temperature, it would take longer than the lifetime of the Universe to find an atom energetic enough to pierce a defect-free membrane of any realistic size. These expectations agree with experiments that have reported no detectable gas permeation through mechanically exfoliated graphene. The highest sensitivity was achieved using micrometre-size wells etched in oxidized silicon wafers, which were sealed with graphene^{1,9,10}. In those measurements, a pressurized gas (for example, helium) could permeate along the SiO₂ layer and gradually fill the microcontainers, making so-called nanoballoons. Their consecutive deflation in air was monitored using atomic force microscopy (AFM) and it was shown that the leakage occurred along only the SiO₂ surface, within minutes but independently of the number of graphene layers used for the sealing¹. These studies allowed the conclusion that graphene membranes are impermeable to all gases, at least with the achieved accuracy of 10^5 – 10^6 atoms per second. This was further corroborated by creating individual atomic-scale defects in graphene nanoballoons, which resulted in their relatively fast deflation/inflation and confirmed the exceptionally high sensitivity of the method^{9,10}.

The devices used in this study were micrometre-size containers made from monocrystals of graphite or hexagonal boron nitride (hBN) using electron-beam lithography and dry etching (Fig. 1, Extended Data

Fig. 1). The containers were sealed with graphene monolayer crystals obtained by mechanical exfoliation and transferred on top of the wells using van der Waals assembly ('Device fabrication' in Methods). In control experiments, bilayer graphene and monolayer molybdenum disulfide (MoS₂) were used for the sealing (see further below and in Methods). The wells were chosen to have an inner diameter d of 0.5 or 1.0 μm , and their depth h was about 50 nm to minimize the containers' volume and, therefore, maximize the sensitivity with respect to the number of inflowing gas molecules. The depth could not be reduced further because van der Waals attraction of graphene to the inner walls caused it to sag^{13,14}, typically by a few tens of nanometres (Fig. 1c, d). The wells' ring-shaped top was typically 1- μm wide to provide a sufficiently large atomically flat area so that no gas diffusion could occur along the resulting 'atomically tight' sealing with its clean and atomically sharp interface^{15,16}. The monocrystalline walls of our microcontainers were also impermeable, as reported previously¹⁷ and confirmed in the present work using wells with walls of different thicknesses. The rough surface outside the wells (due to etching) helped to pin the membranes, preventing their slippage. The atomically tight sealing is the principal difference with respect to the previous experimental setup^{1,9,10} that used 'leaky' SiO₂. In our design, the only possible route for the gas ingress/escape is through the two-dimensional (2D) membrane.

The basic principle used for detection of molecular penetration through graphene membranes is similar to that introduced in ref. ¹ and illustrated in Fig. 1a. The described microcontainers were placed inside a chosen gas atmosphere (for example, helium) and, if graphene

¹Department of Physics and Astronomy, University of Manchester, Manchester, UK. ²National Graphene Institute, University of Manchester, Manchester, UK. ³Key Laboratory of Artificial Micro- and Nano-structures of Ministry of Education, School of Physics and Technology, Wuhan University, Wuhan, China. ⁴Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands. ⁵Chinese Academy of Sciences Key Laboratory of Mechanical Behavior and Design of Materials, Department of Modern Mechanics, University of Science and Technology of China, Hefei, China. ✉e-mail: s.yuan@whu.edu.cn; geim@manchester.ac.uk

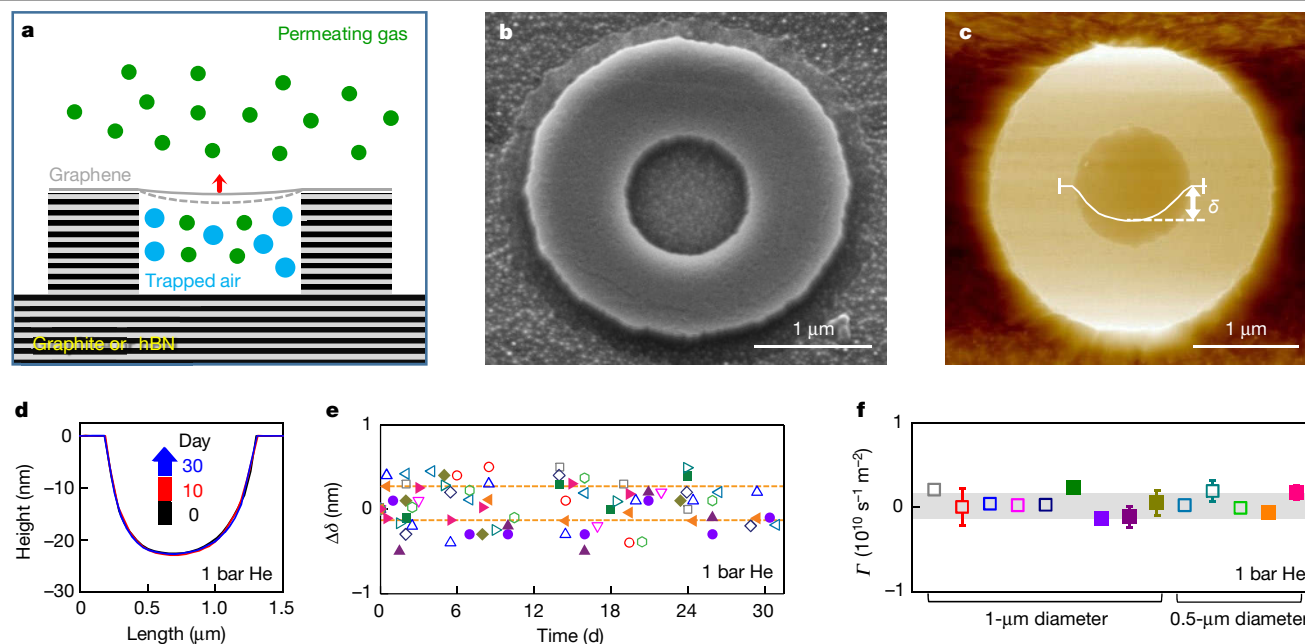


Fig. 1 | The impermeability of graphene to helium. **a**, Schematic of our experimental setup. **b**, Electron micrograph of one of the studied containers. The image was taken at a tilt angle of 20° for a better view. The graphene membrane is seen to stretch over the outer wall and attach to the dry-etched surface outside. **c**, AFM image of a similar device. The white curve shows the profile of the suspended graphene along the well's diameter. The vertical bars indicate the width (about 150 nm) over which such profiles were averaged. **d**, Examples of AFM profiles for the same container after storing it in helium for

days (colour coded). **e**, Changes in the maximum deflection point for 14 containers placed in helium over a one-month period. Different symbols denote containers made from graphite (empty symbols) and hBN (solid). The orange lines indicate the experimental scatter (full range of observed $\Delta\delta$) for one of the devices represented by the same colour. **f**, Permeation rates Γ evaluated from the data in **e**; same symbol coding. Error bars are standard deviation by fitting δ with the linear time dependence. The grey area indicates the overall standard deviation using the data for all the devices.

were permeable to it, the partial pressures inside and outside should equalize so that the total pressure inside the containers (filled with the less permeating air) would increase with time, resulting in gradual lifting and eventual bulging of the membranes. Changes in the membrane position were monitored with AFM. All the containers were first checked for any possible damage to their sealing and the absence of atomic-scale defects^{9,10} as described in Methods ('Experimental procedures'). Only the containers that successfully passed the initial tests were investigated further. They were placed in helium, initially for a few days. After that, the devices were taken out, measured by AFM within 10 min to detect minute changes in the membrane position δ (Fig. 1c) and quickly placed back for further exposure to helium. To maximize our accuracy, AFM mapping was carried out using the PeakForce mode and, for repeated measurements, scans were taken in the same direction over the same area ('AFM measurements' in Methods). For the same reason, we minimized the stress imposed by different pressures inside and outside the containers by normally keeping the external pressure P_a at 1 bar and varying only the partial pressure P of the tested gas (Methods). Furthermore, we avoided containers with d larger than $1\ \mu\text{m}$ because their scans appeared notably noisier ('AFM measurements' in Methods). Our experiments were limited to temperatures $T \leq 60^\circ\text{C}$ because, after thermal cycling to higher T , graphene membranes were often destroyed, probably because of strain induced by thermal expansion/contraction (Extended Data Fig. 2).

Under exposure to helium, no changes in δ could be detected, as detailed in Fig. 1d–f, Extended Data Fig. 3. The figures show that the membrane positions did not change regardless of how long they were exposed to helium. For example, Fig. 1e plots our results for more than a dozen containers over an observation period of one month. None of the devices showed any discernible changes ($\Delta\delta$) in the membranes' original positions, beyond small random fluctuations that did not exceed 0.5 nm in amplitude. Statistical analysis of the $\Delta\delta$ values yielded

a standard deviation of about $1\ \text{\AA}$ (Fig. 1e; 'AFM measurements' in Methods). In control experiments, we carried out the same measurements in air and found fluctuations of a similar amplitude (Extended Data Fig. 3). For higher applied P , the fluctuations were slightly stronger, presumably because the extra pressure caused creep of the membranes (Extended Data Fig. 3).

For small changes $\Delta\delta$ in the membrane position, the number of atoms or molecules ΔN penetrating through the area S is given by^{9,18}

$$\Delta N = c \frac{P_a}{k_B T} S \Delta\delta \quad (1)$$

where $k_B T$ is the thermal energy, k_B is the Boltzmann constant and $c \approx 0.5$ is the coefficient that accounts for the membrane's curved profile ('Evaluation of permeation rates and their accuracy' in Methods). The above accuracy of roughly $1\ \text{\AA}$ over one month translates into no more than a few atoms entering the microcontainers per hour. This accuracy is more than eight orders of magnitude higher than that achieved in the earlier experiments reporting graphene's impermeability^{1,9,10}, which were in turn a few orders of magnitude more sensitive than the detection limit of modern helium leak detectors. In terms of the areal permeation rates $\Gamma = (d\Delta N/dt)/S$, our experiments yield an upper bound of about $10^9\ \text{s}^{-1}\ \text{m}^{-2}$ for possible helium transparency of defect-free graphene. To put this into perspective, monolayer graphene is less permeable than 1-km-thick quartz glass. Furthermore, the found limit allows a lower-bound estimate for the energy barrier E that graphene presents for helium atoms. Using the expression ('Energy barriers' in Methods)

$$\Gamma = \frac{P}{\sqrt{2\pi m k_B T}} \exp\left(-\frac{E}{k_B T}\right) \quad (2)$$

we obtain $E \geq 1.2$ eV, where the pre-exponential factor describes the incident rate of helium atoms and m is their mass. This estimate is consistent with the barriers found theoretically^{2–4}. Keeping in mind that it is hardly possible to improve the accuracy for $\Delta\delta$ beyond 1 Å, the fluctuations increase for larger d , and observations longer than a few months and at considerably higher T are impractical, our results probably present the sensitivity limit of the nanoballoon method.

Helium with its small weakly interacting atoms is recognized as the most permeating of all gases. Nonetheless, we tested several other gases (namely, Ne, N₂, O₂, Ar, Kr and Xe) and, as expected, found no discernible permeation. This places practically the same limit on their E . Unexpectedly, monolayer graphene exhibited noticeable transparency with respect to molecular hydrogen, H₂. We first illustrate this observation qualitatively, by showing in Fig. 2a one of our microcontainers before and after its exposure to hydrogen at 50 °C for 3 d. The membrane clearly bulged up, although the same container passed our impermeability tests with respect to both helium and argon at the same T . This observation is striking because even atomic hydrogen, with a diameter smaller than that of helium, is predicted to experience an E of 2.6–4.6 eV for monolayer graphene^{4–6}, leaving aside the fact that dissociation of molecular hydrogen requires about 4.5 eV, which makes the concentration of atomic hydrogen negligible. Molecular hydrogen is expected³ to have even higher E of more than 10 eV. For such high barriers, hydrogen permeation is completely forbidden and, according to equation (2), it should take billions of years for a single hydrogen atom to get inside the container. In another control experiment, we used microcontainers sealed with bilayer graphene and monolayer MoS₂. They exhibited no detectable permeation under multiday exposure to molecular hydrogen at 50 °C (Extended Data Fig. 4).

To quantify the observed hydrogen permeation, we measured changes in δ as a function of time for many devices at room temperature (295 ± 2 K). They exhibited approximately the same inflation rates within scatter of about $\pm 15\%$, as indicated by the dashed lines in Fig. 2b, which yields $\Gamma \approx 2 \times 10^{10} \text{ s}^{-1} \text{ m}^{-2}$. Note that such a minute gas influx is far beyond the detection limit for microcontainers with SiO₂ sealing^{1,9,10}. Furthermore, working in the regime of small linear-in-time $\Delta\delta$ (no bulging as in Fig. 2a), we measured hydrogen permeation at different T . The temperature dependences followed the Arrhenius law, $\Gamma \propto \exp(-E/k_B T)$, yielding an activation barrier of 1.0 ± 0.1 eV (Fig. 2c). This relatively small E strongly disagrees with the theoretical expectations and, more importantly, with the fact that smaller helium atoms did not penetrate through the same membranes.

Trying to understand the origin of the unexpected behaviour, we performed two additional sets of experiments. First, we quantified the hydrogen permeation rates at different pressures P and found $\Gamma \propto P^{1/2}$ (Extended Data Fig. 5). The square-root dependence is characteristic of processes involving an equilibrium between adsorbed and desorbed constituents of a bipartite gas¹⁹, in contrast to the linear dependence of equation (2) valid for weakly interacting atoms ('Energy barriers' in Methods). Second, we measured permeation for hydrogen's isotope deuterium. Within our detection limit, no permeation could be discerned, which puts a limit of $\Gamma \leq 10^9 \text{ s}^{-1} \text{ m}^{-2}$ on the deuterium influx ('Isotope effect' in Methods and Extended Data Fig. 8).

To understand the reason for the exclusivity of hydrogen among the other gases, let us recall the following facts. Locally curved and strained graphene surfaces are known experimentally to be chemically reactive^{20,21} and are expected to lower the energy required for dissociation of molecular hydrogen^{22,23}. For a local protrusion (ripple) with $t/D \geq 5\%$ (where t is its height and D the lateral size), the dissociated state with two hydrogen adatoms becomes energetically more favourable^{22,23}, whereas the energy barrier required to reach this state is also reduced to about 1 eV ('Ab initio simulations of graphene's catalytic activity' in Methods). This catalytic activity of graphene is relevant because suspended membranes exhibit extensive nanoscale rippling^{24–26} with t/D that can easily exceed 5% for both static^{24,25} and dynamic²⁶ ripples

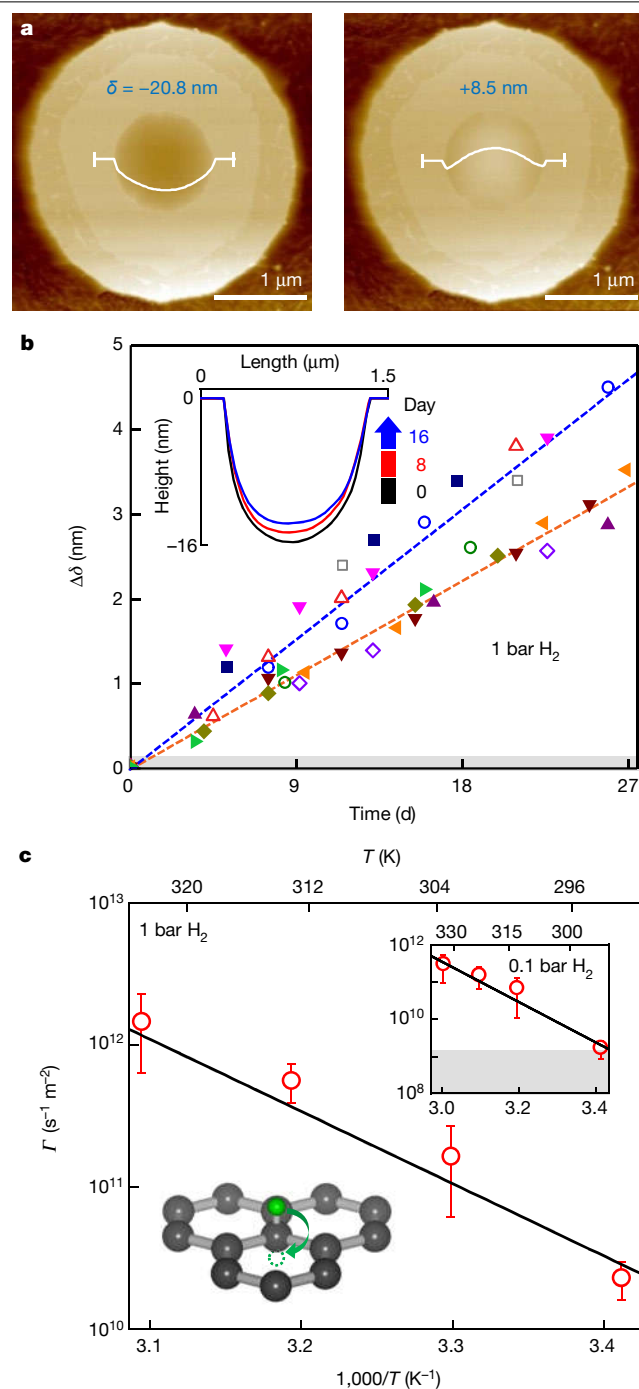


Fig. 2 | Hydrogen permeation through defect-free graphene. **a**, AFM micrographs of the same microcontainer before (left) and after (right) storing it for 3 d in molecular hydrogen at 1 bar. To speed up permeation, the gas was heated to 50 °C. White curves show the height profiles along the well's diameter. Tens of microcontainers were tested, showing the same effect. The somewhat darker outside region of the well's top appears because it is not atomically flat but slightly tapered (our lithography masks often thinned towards the outside perimeter, allowing some plasma etching of the rim region). **b**, Time evolution of $\Delta\delta$ for 12 different devices in molecular hydrogen at 1 bar at $T = 295 \pm 2$ K. The empty and solid symbols denote graphite and hBN wells, respectively. Blue and orange dashed lines are the best linear fits for two of the devices (colour coded) to indicate experimental scatter. Inset: representative changes in the AFM profiles with time. **c**, Hydrogen permeation rates at different T . Symbols are experimental data and the solid curve is the best fit to the activation behaviour, which yields $E = 1.0 \pm 0.1$ eV. Top inset: same as the main panel, but for $P = 0.1$ bar. Error bars are standard deviation using six or more devices for each T . Bottom inset: illustration of the flipping process in the suggested mechanism of hydrogen permeation. The grey areas in **b**, **c** indicate our detection limit.

(Extended Data Fig. 7). Another fact is that monolayer graphene is known to be highly permeable to protons, exhibiting an activation energy of 1.0 ± 0.05 eV, whereas bilayer graphene and monolayer MoS₂ exhibit no detectable proton permeation^{11,12}. This is relevant because a hydrogen atom adsorbed on graphene shares its electron with the conducting surface and is indistinguishable from an adsorbed proton. Furthermore, it is also known that deuterons, nuclei of deuterium atoms, experience a higher barrier than protons, which drastically slows their permeation through monolayer graphene¹² ('Isotope effect' in Methods).

On the basis of the above facts, we propose the following scenario for the observed hydrogen permeation. First, molecular hydrogen is chemisorbed (adsorbed and dissociated) on graphene ripples, which results in *sp*³-bonded adatoms as illustrated in Extended Data Fig. 6. These adatoms then flip to the other side of graphene in a 1.0-eV transfer process similar to that previously reported for proton transport^{11,12} (inset of Fig. 2c). The flipped adatoms subsequently desorb from the concave surface. This scenario is fully consistent with all the experimental evidence and, also, explains why the observed permeation is limited to hydrogen and monolayer graphene. Indeed, among the tested 2D crystals, only the latter is sufficiently transparent to protons. Neither bilayer graphene nor monolayer MoS₂ allow protons through¹¹, whereas monolayer graphene presents a notably higher barrier for heavier deuterons than protons¹² (see 'Isotope effect' in Methods).

Although our experiments cannot distinguish directly whether it is chemisorption or flipping that limits the hydrogen permeation, the close match of the measured *E* with the value reported in ref.¹² hints that the flipping is likely to be the rate-limiting process. This is also supported by the observed isotope effect. Indeed, our DFT calculations could not find any influence of zero-point oscillations on hydrogen's dissociation (see 'Isotope effect' in Methods). However, the flipping is expected to exhibit an isotope shift because zero-point oscillations decrease the energy of the initial state in the transfer process¹². This shift results in a higher effective barrier for deuterium and makes its permeation undetectable in our experiments (see 'Isotope effect' in Methods). The dependence $\Gamma \propto P^{1/2}$ (suggesting a finite coverage of graphene with hydrogen) is also consistent with the flipping step being the limiting factor. Indeed, it is easier for lighter adatoms to desorb from graphene (because of stronger zero-point oscillations), which should result in higher coverage of the graphene surface with deuterium. Accordingly, if chemisorption were the limiting step, higher permeation rates would be expected for deuterium rather than hydrogen, contrary to our observations.

To conclude, defect-free graphene should be completely impermeable to all atomic and molecular species at room temperature, but ripples, wrinkles and other defects inducing a local curvature are catalytically active and allow non-negligible permeation of hydrogen. If necessary, the latter can be blocked by using bilayer graphene or other 2D materials such as monolayer MoS₂. Our results have implications for many observations in the literature. For example, ripples are likely to play an important role in lowering barriers for proton transport through 2D membranes^{11,12}, a distinct possibility not considered so far theoretically^{7,27–29}. Similarly, the observations may shed light on the intercalation of graphene on silicon carbide by molecular hydrogen that is argued to permeate through defect-free graphene^{30,31}. The discussed processes may also be critical for the interaction of graphene with water and hydrocarbons and, more generally, emphasize unexpectedly high catalytic activity of non-flat graphene, in stark contrast to the extreme chemical inertness of its bulk counterpart, graphite.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2070-x>.

- Bunch, J. S. et al. Impermeable atomic membranes from graphene sheets. *Nano Lett.* **8**, 2458–2462 (2008).
- Berry, V. Impermeability of graphene and its applications. *Carbon* **62**, 1–10 (2013).
- Leenaerts, O., Partoens, B. & Peeters, F. M. Graphene: a perfect nanoballoon. *Appl. Phys. Lett.* **93**, 193107 (2008).
- Tsetseris, L. & Pantelides, S. T. Graphene: an impermeable or selectively permeable membrane for atomic species? *Carbon* **67**, 58–63 (2014).
- Miao, M., Nardelli, M. B., Wang, Q. & Liu, Y. First principles study of the permeability of graphene to hydrogen atoms. *Phys. Chem. Chem. Phys.* **15**, 16132–16137 (2013).
- Seel, M. & Pandey, R. Proton and hydrogen transport through two-dimensional monolayers. *2D Mater.* **3**, 025004 (2016).
- Feng, Y. et al. Hydrogenation facilitates proton transfer through two-dimensional honeycomb crystals. *J. Phys. Chem. Lett.* **8**, 6009–6014 (2017).
- Wang, W. L. & Kaxiras, E. Graphene hydrate: theoretical prediction of a new insulating form of graphene. *New J. Phys.* **12**, 125012 (2010).
- Koenig, S. P., Wang, L., Pellegrino, J. & Bunch, J. S. Selective molecular sieving through porous graphene. *Nat. Nanotechnol.* **7**, 728–732 (2012).
- Wang, L. et al. Molecular valves for controlling gas phase transport made from discrete ångström-sized pores in graphene. *Nat. Nanotechnol.* **10**, 785–790 (2015).
- Hu, S. et al. Proton transport through one-atom-thick crystals. *Nature* **516**, 227–230 (2014).
- Lozada-Hidalgo, M. et al. Sieving hydrogen isotopes through two-dimensional crystals. *Science* **351**, 68–70 (2016).
- Bunch, J. S. et al. Electromechanical resonators from graphene sheets. *Science* **315**, 490–493 (2007).
- Radha, B. et al. Molecular transport through capillaries made with atomic-scale precision. *Nature* **538**, 222–225 (2016).
- Haigh, S. J. et al. Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. *Nat. Mater.* **11**, 764–767 (2012).
- Kelly, D. J. et al. Nanometer resolution elemental mapping in graphene-based TEM liquid cells. *Nano Lett.* **18**, 1168–1174 (2018).
- Hu, S. et al. Transport of hydrogen isotopes through interlayer spacing in van der Waals crystals. *Nat. Nanotechnol.* **13**, 468–472 (2018).
- Koenig, S. P., Boddeti, N. G., Dunn, M. L. & Bunch, J. S. Ultrastrong adhesion of graphene membranes. *Nat. Nanotechnol.* **6**, 543–546 (2011).
- Deveau, N. D., Ma, Y. H. & Datta, R. Beyond Sieverts' law: a comprehensive microkinetic model of hydrogen permeation in dense metal membranes. *J. Membr. Sci.* **437**, 298–311 (2013).
- Wu, Q. et al. Selective surface functionalization at regions of high local curvature in graphene. *Chem. Commun.* **49**, 677–679 (2013).
- Bissett, M. A., Konabe, S., Okada, S., Tsuji, M. & Ago, H. Enhanced chemical reactivity of graphene induced by mechanical strain. *ACS Nano* **7**, 10335–10343 (2013).
- Boukhalov, D. W. & Katsnelson, M. I. Enhancement of chemical activity in corrugated graphene. *J. Phys. Chem. C* **113**, 14176–14178 (2009).
- McKay, H., Wales, D. J., Jenkins, S. J., Verges, J. A. & de Andres, P. L. Hydrogen on graphene under stress: molecular dissociation and gap opening. *Phys. Rev. B* **81**, 075425 (2010).
- Meyer, J. C. et al. On the roughness of single- and bi-layer graphene membranes. *Solid State Commun.* **143**, 101–109 (2007).
- Geringer, V. et al. Intrinsic and extrinsic corrugation of monolayer graphene deposited on SiO₂. *Phys. Rev. Lett.* **102**, 076102 (2009).
- Fasolino, A., Los, J. H. & Katsnelson, M. I. Intrinsic ripples in graphene. *Nat. Mater.* **6**, 858–861 (2007).
- Kroes, J. M. H., Fasolino, A. & Katsnelson, M. I. Density functional based simulations of proton permeation of graphene and hexagonal boron nitride. *Phys. Chem. Chem. Phys.* **19**, 5813–5817 (2017).
- Poltavsky, I., Zheng, L., Mortazavi, M. & Tkatchenko, A. Quantum tunneling of thermal protons through pristine graphene. *J. Chem. Phys.* **148**, 204707 (2018).
- Mazzuca, J. W. & Haut, N. K. Theoretical description of quantum mechanical permeation of graphene membranes by charged hydrogen isotopes. *J. Chem. Phys.* **148**, 224301 (2018).
- Riedl, C., Coletti, C., Iwasaki, T., Zakharov, A. A. & Starke, U. Quasi-free-standing epitaxial graphene on SiC obtained by hydrogen intercalation. *Phys. Rev. Lett.* **103**, 246804 (2009).
- Kunc, J., Rejthorn, M. & Hlídek, P. Hydrogen intercalation of epitaxial graphene and buffer layer probed by mid-infrared absorption and Raman spectroscopy. *AIP Adv.* **8**, 045015 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Crown 2020

Methods

Device fabrication

Our devices were fabricated as shown schematically in Extended Data Fig. 1a. A monocrystal of either graphite (NGS Naturgraphit) or hBN (HQ Graphene) with a thickness of at least 150 nm was first mechanically exfoliated onto an oxidized silicon wafer that was freshly cleaned in an oxygen plasma. The quality of the crystal's top surface was carefully checked for the presence of atomic terraces using dark-field and differential-interference-contrast microscopy. These modes allow detection of crystal edges, terraces and tears, even for a monolayer thickness. Using e-beam lithography, a set of ring-shaped polymer masks with an inner diameter d of 0.5 or 1.0 μm was patterned on atomically flat parts of the surface (without terraces). Reactive-ion etching was then used to remove roughly 50 nm of the exposed area to form micrometre-diameter wells (Extended Data Fig. 1b). After the lithography mask was dissolved, we annealed the structures in a hydrogen and argon (1:10) atmosphere at 400 $^{\circ}\text{C}$ for 6 h. Then a relatively large crystal of monolayer graphene (also obtained by mechanical exfoliation) was transferred in air on top of the wells using the standard transfer procedures for assembly of van der Waals heterostructures^{32,33}. Extended Data Fig. 1b shows an optical image of an array of graphene-sealed hBN wells. Closer views of such microcontainers are provided in Fig. 1, Extended Data Fig. 2a, where one can clearly see graphene membranes draping over the wells. For comparison, Extended Data Fig. 2b shows a broken graphene membrane after our unsuccessful attempt to test it at 80 $^{\circ}\text{C}$. As the inner walls of the containers are not perfectly round (see, for example, Fig. 1b, c), graphene membranes sag inside in a slightly asymmetric manner as noticeable in some AFM profiles (for example, Fig. 1d).

Experimental procedures

After the fabrication, microcontainers were first checked with AFM for possible tears, wrinkles and other defects. Only devices with seemingly perfect sealing were used for further investigation. Those were tested further by placing them in a 3-bar argon atmosphere overnight. Occasionally, we found inflated containers that deflated quickly in air. In principle, this could be due to defects^{9,10,34}, but in most cases we could trace the leakage to poor sealing of the microcontainers: either the top surface of the wells was slightly damaged by dry etching so that the rough streaks connected the inner and outer rim edges or small wrinkles were present, as retrospectively revealed by dedicated AFM analysis and scanning electron microscopy. The devices that passed the above tests were placed in a tested gas atmosphere and their possible inflation was carefully monitored as described in the main text.

For gas tests, microcontainers were placed inside a small stainless-steel chamber. It was evacuated to approximately 10^{-3} mbar and then one of the gases under investigation was introduced inside. For studies of temperature dependences, the whole vacuum chamber was placed inside an oven with controllable T .

AFM measurements

To monitor changes in the position of the graphene membranes, we employed the PeakForce mapping mode (Dimension FastScan from Bruker). The use of this AFM mode was essential to achieve the highest possible accuracy in our measurements of the membrane position. The PeakForce mode minimizes tip-to-sample interactions by employing an imaging force that can be as small as about 1 nN and has little effect on suspended graphene. For comparison, when we tried the contact-mode AFM imaging, graphene membranes were found to sag down after each scan by as much as several nanometres, which was obviously unacceptable for our purposes. The PeakForce mode also allows straightforward analysis of the obtained scans, compared with the tapping AFM mode where a non-negligible pressure induced by the tip requires rather involved deconvolution of AFM images^{1,35}. Furthermore, to maximize reproducibility between consecutive PeakForce

scans, they were taken always along the same direction, x , across the centre of suspended graphene membranes and averaged over a finite width of roughly 150 nm (vertical bars in the shown AFM images). This approach also allowed us to increase accuracy by avoiding changes in AFM profiles caused by the slight asymmetry in membranes' sagging, as pointed out in the Methods section 'Device fabrication'. The asymmetry remains constant during measurements for a given device and does not affect our results. The averaged deflection profile $\delta(x)$ allowed us to detect minor changes $\Delta\delta$ in a membrane's lowest position at the well's centre (denoted above as δ) such that $\delta = \delta(0) + \Delta\delta$, where $\delta(0)$ is the initial-in-time position at the well's centre.

Extended Data Fig. 3a, b shows the accuracy and reproducibility of our measurements of $\delta(x)$ and $\Delta\delta$. For this dataset, two microcontainers were measured in air as described above, and ten AFM scans were taken at 1-h intervals. Between each scan, the devices were taken out of the AFM setup and then placed back to mimic the real measurement procedures. The figures show that the profiles $\delta(x)$ were very stable, and the resulting $\Delta\delta$ did not exceed about 0.3 nm. The statistical uncertainty (standard deviation) for this set of AFM measurements was about 0.16 nm. To check for longer-term stability, 12 microcontainers with $d = 0.5$ and 1.0 μm and different sagging (maximum depth $\delta(0)$ varied between 5 and 25 nm) were kept in air for more than 20 d. Their height profiles were captured at regular intervals, a few days apart. The devices also exhibited excellent stability such that $\Delta\delta$ did not exceed 0.3 nm (Extended Data Fig. 3c), in good agreement with the short-term results in Extended Data Fig. 3a.

After exposing microcontainers to higher helium pressures, we again did not observe any discernible changes in $\delta(x)$ but random fluctuations in $\Delta\delta$ increased (Extended Data Fig. 3e, f), presumably because of additional stresses induced by pressure. Note that, trying to improve our measurement accuracy further, we also made and tested microcontainers with $d \geq 2 \mu\text{m}$. However, their stability was much worse, with $\Delta\delta$ exceeding 1 nm, probably due to increasing instabilities caused by tip-membrane interactions. Data from such wells were not used in the reported analysis.

Evaluation of permeation rates and their accuracy

Equation (1) can be deduced from the expression derived in ref. ⁹ as follows. The pressure P inside the container includes two components: one is the initial atmospheric pressure of the trapped air (P_a) and the other is ΔP , the pressure change induced by molecular permeation. For small changes in the membrane position, ΔP can be estimated from the Henry solution³⁶ as

$$\Delta P = K(v)EL(\Delta\delta)^3/a^4 \quad (3)$$

where E is Young's modulus, L is the membrane thickness, a is the radius of the container and $K(v)$ is the coefficient that depends on Poisson's ratio v . For graphene^{18,37}, $E = 1 \text{ TPa}$, $v = 0.16$ and $K(v) = 3.09$. The gas volume inside the container is given by $V = V_0 + \Delta V = Sh + cS\delta$, where S and h are the container's area and depth, respectively, and $c \approx 0.5$ is the numerical coefficient accounting for the curved shape of the membrane. Putting the above expressions for P and V into the ideal gas law $PV = (N_0 + \Delta N)k_B T$ (N_0 is the initial number of air molecules), we find

$$\Delta N = (P_a \Delta V + V_0 \Delta P + \Delta P \Delta V)/k_B T \quad (4)$$

Substituting the expressions for ΔV and ΔP , equation (4) can be written as⁹

$$\Delta N = \frac{S}{k_B T} \left[cP_a \Delta\delta + (h + c\delta(0)) \frac{K(v)EL}{a^4} (\Delta\delta)^3 + c \frac{K(v)EL}{a^4} (\Delta\delta)^4 \right] \quad (5)$$

For the known constants and noticing that the largest deflection $\Delta\delta$ used in our quantitative analysis was only about 4 nm, we find that the

Article

second and third terms of equation (5) should not exceed 20% of the linear term's value. Therefore, for the purpose of our analysis, equation (5) can be simplified to equation (1). This also agrees with the fact that $\Delta\delta$ evolved linearly in time, within our experimental scatter (Fig. 2b). If contributions of the nonlinear terms were considerable, $\Delta\delta$ should start evolving nonlinearly as a function of ΔN and, hence, time.

Energy barriers

Helium permeation through the barrier presented by a graphene membrane can be estimated using

$$\Gamma = A \exp\left(-\frac{E}{k_B T}\right) \quad (6)$$

where E is the energy barrier for incident atoms and A is their attempt rate (that is, the number of atoms striking a unit area per second). Weakly interacting helium atoms are not adsorbed onto graphene and, therefore, the attempt rate is given by³⁸

$$A = \frac{1}{4} \frac{N}{V} v_G = \frac{P}{4 k_B T} v_G \quad (7)$$

where $v_G = \sqrt{8k_B T / \pi m}$ is the mean speed of helium atoms and m is their atomic weight. Combining equations (6) and (7), we obtain equation (2). However, if gas atoms or molecules become adsorbed on a graphene surface, like in the case of hydrogen, equation (7) is no longer applicable, and the attempt rate depends on an equilibrium density of adsorbed species. Under the latter circumstances, the different dependence $A \propto \sqrt{P}$ is expected¹⁹, in agreement with our results for hydrogen in Extended Data Fig. 5.

Ab initio simulations of graphene's catalytic activity

Energy barriers for the dissociation of molecular hydrogen on flat and rippled graphene were calculated from first principles using DFT, as implemented in the Vienna ab initio package³⁹. The generalized gradient approximation⁴⁰ and projected augmented wave were adopted to describe the exchange correlation potential and ion–electron interactions. The kinetic energy cutoff and k -point mesh were set to 500 eV and $7 \times 7 \times 1$, respectively⁴¹. A vacuum region of 20 Å was used to avoid the periodic interaction. The stress force and energy convergence criteria were chosen as 0.01 eV Å⁻¹ and 10⁻⁵ eV, respectively. The van der Waals interactions were included in the dissociation process and treated by the semi-empirical DFT-D3 method^{42,43}. A supercell of 8×8 graphene unit cells was adopted for the simulations, and ripples were characterized by the ratio t/D of their height t to the corrugation diameter D (inset in Extended Data Fig. 6a). The energy barrier for the reaction pathway was calculated using the climbing-image nudged elastic band method, in which the total energies of initial, final and several intermediate states during the reaction process were calculated explicitly⁴⁴. The initial state was constructed as follows²². First, we created a corrugated graphene supercell with a certain t/D by allowing the atomic structure to relax under biaxial compression. Next, two hydrogen atoms were attached to specified carbon atoms, and the whole system was allowed to relax to its ground state, during which the positions of unoccupied carbon atoms were fixed to keep the t/D value constant. The relaxed carbon structure was then used as the initial configuration and the electron distribution was optimized during the reaction process.

For a given t/D , there are many possible corrugated configurations. If we consider high-symmetry configurations, the corrugation centre is located either at the top of a carbon atom or between two nearest neighbours or at the hexagon centre. To minimize the dissociation energy, we relaxed the above three structures of rippled graphene with two adsorbed hydrogen atoms and used them as the initial states before hydrogenation. The initialized graphene ripple could be allowed to relax further before chemical reaction, but we found that this caused

little effect on the energy barrier. After trying many different configurations and reaction processes, we found that the dissociation energy reached a minimum when two opposed sites in a hexagon were hydrogenated (see the insets in Extended Data Fig. 6a). In Extended Data Fig. 6a, we show changes in the total energy during the reaction process for $t/D = 7.5\%$. The dissociation energy barrier is about 1.1 eV and given by the difference between the initial and highest energy states along the reaction pathway. For comparison, the dissociation energy of molecular hydrogen in vacuum is about 4.5 eV (ref. ⁴⁵), which shows that ripples are highly catalytically active.

The dissociation energy depends on where in the unit cell hydrogen atoms are adsorbed. For example, Extended Data Fig. 6b shows the adsorption process for the same t/D as in Extended Data Fig. 6a but with hydrogen atoms attached to the nearest carbon atoms. In this case, the dissociation energy barrier is higher (approximately 2.9 eV). Our results for different ripple curvatures t/D are plotted in Extended Data Fig. 6c. Clearly, the dissociation energy decreases monotonically with increasing curvature, and the changes become rather gradual for $t/D > 4\%$. Note that the critical curvature, at which ripples become energetically favourable for dissociation of molecular hydrogen ($t/D \approx 2.5\%$), is smaller than $t/D \approx 4\%$, which was reported in the earlier study²². This is because of improvements in the simulation method and optimized atomistic configurations.

Although graphene membranes are known to contain numerous extrinsic (static) ripples^{24,25} that have typical $t/D \approx 5\%$, it is instructive to find what kind of intrinsic (dynamic) ripples one can expect due to thermal fluctuations²⁶. To this end, we performed molecular dynamics simulations using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)⁴⁶ and graphene membranes consisting of 387,200 atoms. Periodic boundary conditions were usually employed to mimic an infinite membrane, but we also performed simulations for finite-size membranes (from about 35 to 100 nm in diameter). The lateral size of ripples (D) ranged between a few and 10 nm, independently of the membrane size. Their typical configurations at different T were obtained after thermalization in 100,000 steps (0.00025 femtoseconds per step) and averaging over 20 of such snapshots (Extended Data Fig. 7a). Extended Data Fig. 7b, c shows the areal density for ripples with $t/D \geq 4\%$, which are most catalytically active. One can see that thermal fluctuations generate many such ripples that can result in dissociation of molecular hydrogen.

It is not clear whether static or dynamic ripples dominate the adsorption–dissociation process for graphene membranes. One of the issues limiting a contribution from thermally excited ripples could be their relatively short lifetimes. Our simulations show that the mean half-life of a ripple, during which its t/D drops to half, is of the order of femtoseconds. For comparison, permeation of adsorbed hydrogen atoms through graphene involves the timescale τ , which can be estimated from their adsorption energy as $E_{ad} = h/\tau$ where h is Planck's constant. For atomic hydrogen on graphene^{47,48}, E_{ad} is expected to be approximately 0.4–1.0 eV, and this yields τ of a few femtoseconds. This is of the same order of magnitude as the characteristic lifetime of ripples.

Besides static and dynamic ripples, there are strained areas around the rim of our microcontainers, which in principle could also contribute to the observed hydrogen permeation. However, this scenario is ruled out by the experimental fact that the observed permeability was proportional to the membrane's area rather than its circumference. Furthermore, permeation rates were the same for devices with different sagging (varying from about 5 to 40 nm), which led to different strain. The above conclusion is also supported by our DFT calculations in which the chemisorption process was considered for flat graphene under strain. Extended Data Fig. 6d shows that the dissociation energy remains high (about 3 eV) even for strains as high as about 15%. This proves that it is the curvature rather than strain that is important for the chemisorption process.

Isotope effect

The proposed mechanism of hydrogen permeation involves several steps: dissociation of molecular hydrogen on graphene ripples leading to a finite coverage of the surface with hydrogen atoms; flipping of the atoms across the membrane in a proton-like transfer process; and their recombination and desorption as molecular hydrogen. Only the chemisorption and flipping are expected to involve sufficiently high energy barriers as discussed in the main text. To determine which of these two key barriers limits the observed permeation, we performed experiments using deuterium. To this end, ten microcontainers were sealed with monolayer graphene and exposed to deuterium at 1 bar at room temperature. $\Delta\delta$ was monitored as a function of time. In stark contrast to our molecular hydrogen experiments, the devices did not show any discernible changes in $\Delta\delta$ (Extended Data Fig. 8). The absence of deuterium permeation was further verified in elevated- T tests such as those in Fig. 2a. Several devices were exposed to deuterium at 50 °C for 3 d, but none exhibited any bulging, in contrast to the molecular hydrogen experiments of Fig. 2a.

To understand the isotope effect, we carried out DFT calculations for chemisorption of deuterium on rippled graphene. After including corrections due to zero-point oscillations, we found little difference in the dissociation energies with respect to molecular hydrogen (Extended Data Fig. 6). However, desorption of deuterium from the graphene surface should be slower because of the same quantum corrections. Hence, surface coverage for deuterium atoms should be higher than that for hydrogen. The latter isotope effect is well known for both graphene and graphite^{47,48} and implies that, if chemisorption were the rate-limiting process, higher permeation rates would have been expected for deuterium rather than molecular hydrogen, contrary to our observations. The latter conclusion indicates again that the limiting process is the proton flipping. Indeed, in this case, one expects another isotope effect analogous to that observed in the transport experiments of refs. 12,49 where deuterons exhibited ten times lower conductivity through graphene than protons. The reduction factor R has been attributed to the fact that deuterons have a lower energy in their initial bound state in Nafion^{12,49}, which results in an effectively higher barrier for flipping (by $\Delta E \approx 60$ meV). This energy shift leads to the ratio between hydrogen and deuterium permeation, which is given by $R = \exp(\Delta E/k_B T) \approx 10$ (ref. 12). For the case of atomic hydrogen/deuterium adsorbed on graphite, ΔE was found⁴⁸ to be around 90 meV. Assuming the same value for graphene, this shift should result in about 35 times slower permeation of deuterons with respect to protons. Because of the detection limit of about $10^9 \text{ s}^{-1} \text{ m}^{-2}$, we can only conclude from our experiments that $R \geq 20$, in good agreement with the above expectations. This supports the proposed scenario in which the flipping step limits hydrogen permeation through monolayer graphene.

Data availability

All the mentioned data to support this study and its conclusions are available upon request from P.Z.S. (pengzhan.sun@manchester.ac.uk).

32. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499**, 419–425 (2013).
33. Wang, L. et al. One-dimensional electrical contact to a two-dimensional material. *Science* **342**, 614–617 (2013).
34. Park, H. G. & Jung, Y. Carbon nanofluidics of rapid water transport for energy applications. *Chem. Soc. Rev.* **43**, 565–576 (2014).
35. Whittaker, J. D., Minot, E. D., Tanenbaum, D. M., McEuen, P. L. & Davis, R. C. Measurement of the adhesion force between carbon nanotubes and a silicon dioxide substrate. *Nano Lett.* **6**, 953–957 (2006).
36. Hencky, H. Über den spannungszustand in kreisrunden platten mit verschwindender biegungssteifigkeit. *Z. Math. Phys.* **63**, 311–317 (1915).
37. Wang, G. et al. Measuring interlayer shear stress in bilayer graphene. *Phys. Rev. Lett.* **119**, 036101 (2017).
38. Landau, L. D. & Lifshitz, E. M. *Course of Theoretical Physics Vol. 5 Statistical Physics* 3rd edn (Pergamon Press, 1980).
39. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
40. Perdew, J. P., Burke, K. & Ernzerhof, M. M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
41. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
42. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
43. Kerber, T., Sierka, M. & Sauer, J. Application of semiempirical long-range dispersion corrections to periodic systems in density functional theory. *J. Comput. Chem.* **29**, 2088–2097 (2008).
44. Sheppard, D., Xiao, P., Chemelewski, W., Johnson, D. D. & Henkelman, G. A generalized solid-state nudged elastic band method. *J. Chem. Phys.* **136**, 074103 (2012).
45. Herzberg, G. & Monfils, A. The dissociation energies of the H_2 , HD, and D_2 molecules. *J. Mol. Spectrosc.* **5**, 482–498 (1961).
46. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
47. Hornekær, L. et al. Clustering of chemisorbed H(D) atoms on the graphite (0001) surface due to preferential sticking. *Phys. Rev. Lett.* **97**, 186102 (2006).
48. Paris, A. et al. Kinetic isotope effect in the hydrogenation and deuteration of graphene. *Adv. Funct. Mater.* **23**, 1628–1635 (2013).
49. Bukola, S. & Creager, S. E. A charge-transfer resistance model and Arrhenius activation analysis for hydrogen ion transmission across single-layer graphene. *Electrochim. Acta* **296**, 1–7 (2019).

Acknowledgements This work was supported by the Lloyd's Register Foundation, the European Research Council (grants ARTIMATTER and VANDER), Graphene Flagship and the Royal Society. S.J.Y. acknowledges support from the National Key R&D Program of China (grant 2018YFA0305800) and Supercomputing Center of Wuhan University.

Author contributions A.K.G. suggested and directed the project with help from P.Z.S., Q.Y. and F.C.W. P.Z.S., Q.Y., W.J.K. and Y.V.S. fabricated the devices, performed measurements and analysed the data. W.Q.X., J.Y., M.I.K., S.J.Y. and F.C.W. provided theoretical support. I.V.G., R.R.N., F.C.W. and M.L.-H. contributed to interpretation of the experimental results. A.K.G., P.Z.S., I.V.G. and M.L.-H. wrote the manuscript. All authors contributed to discussions.

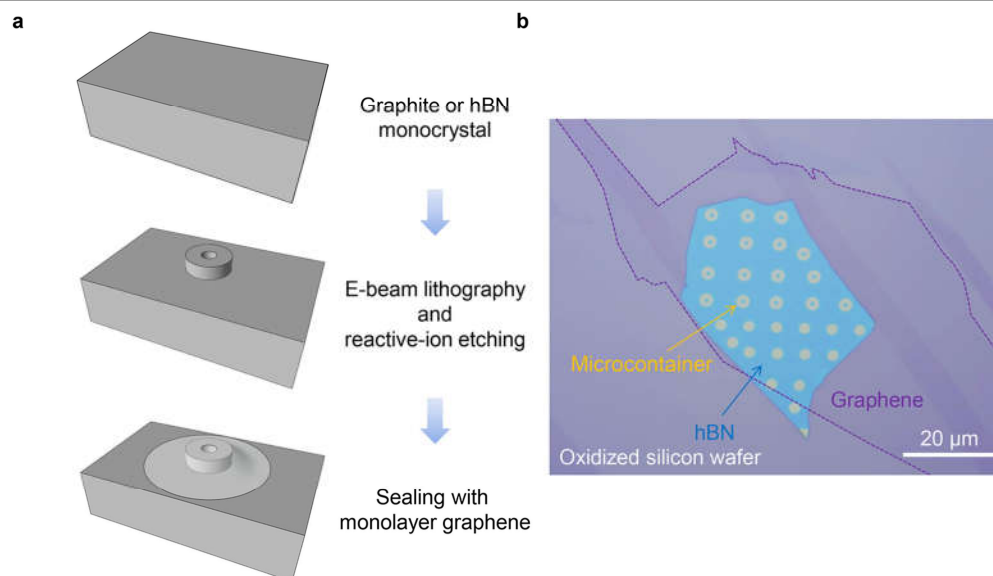
Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.J.Y. or A.K.G.

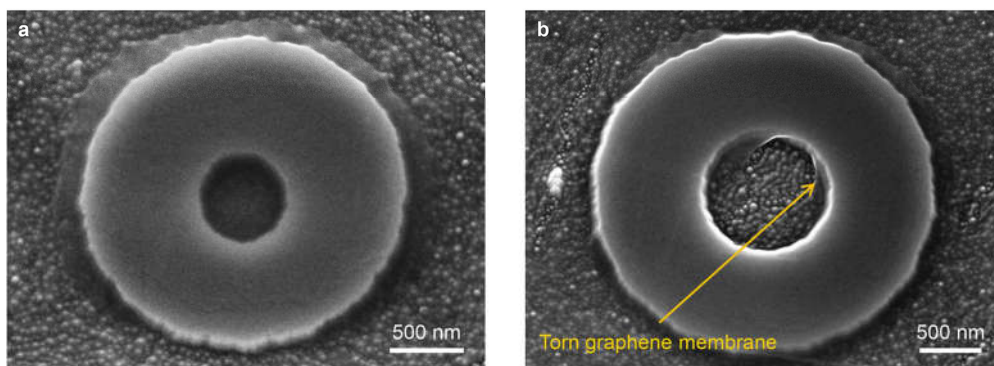
Peer review information *Nature* thanks Rohit Karnik, Valentina Tozzini and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

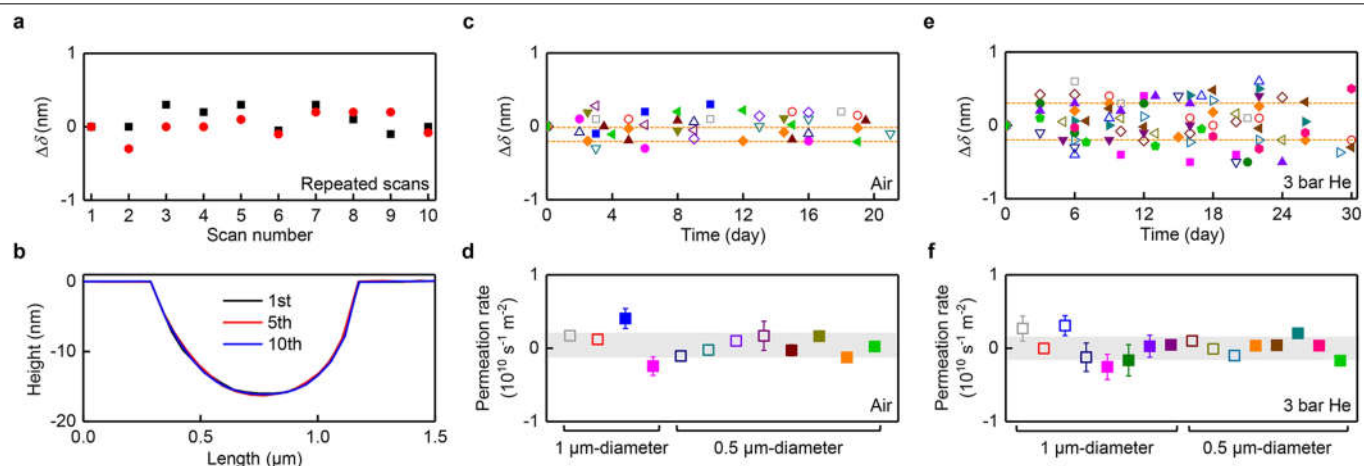


Extended Data Fig. 1 | Device fabrication. **a**, Graphite or hBN monocrystals are obtained by mechanical exfoliation. Micrometre-size wells are then made by e-beam lithography and ion etching. Monolayer graphene is transferred on top

to seal the wells. **b**, Optical micrograph of a set of hBN microcontainers. The dashed curve indicates the position of monolayer graphene.



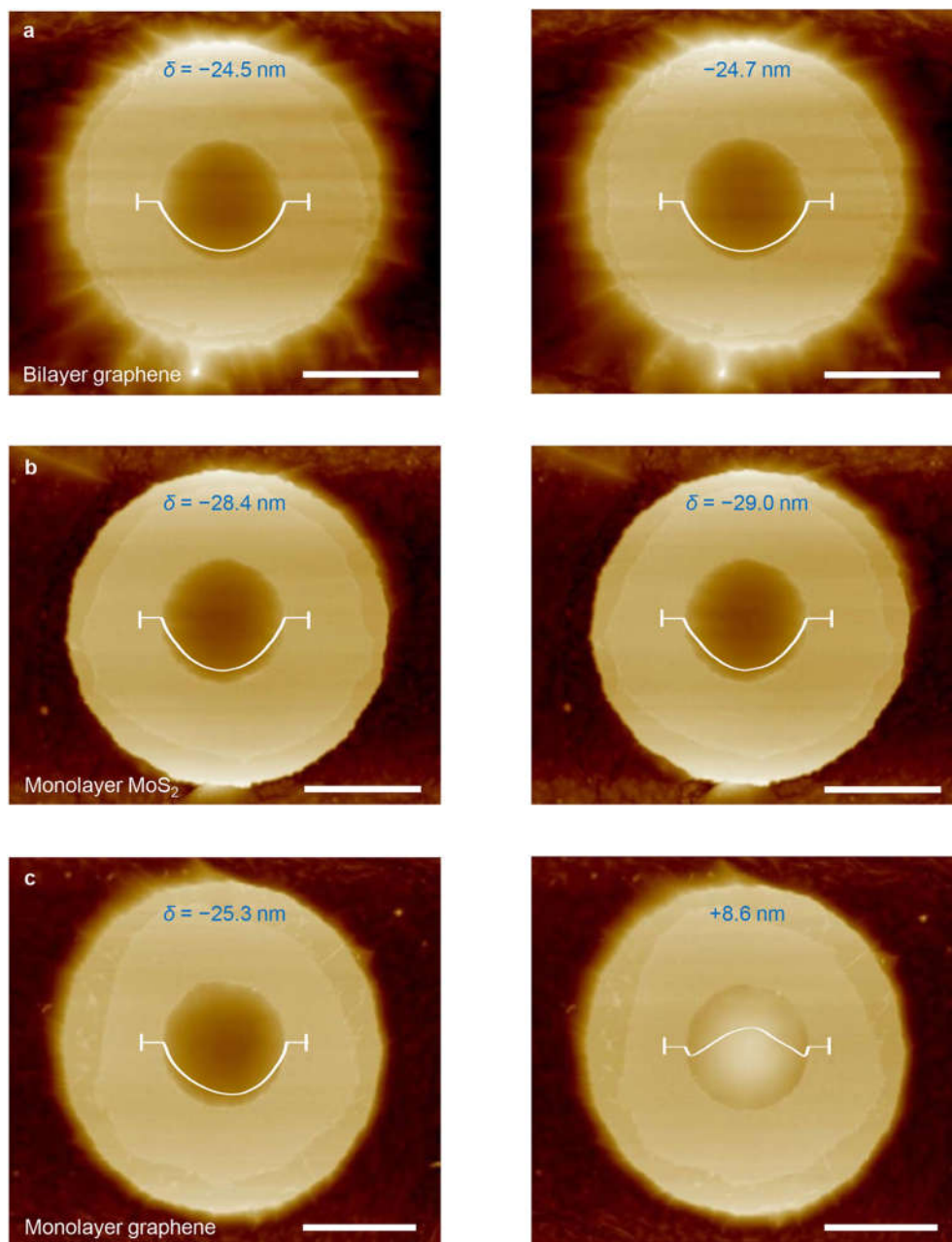
Extended Data Fig. 2 | Electron micrographs of our microcontainers. **a**, A microcontainer with $d = 0.5 \mu\text{m}$. Such images were taken only after finishing measurements to avoid electron-beam damage. **b**, Example of a broken graphene seal: the membrane was damaged after a thermal cycle to 80°C .



Extended Data Fig. 3 | Stability of graphene membranes in air and helium.

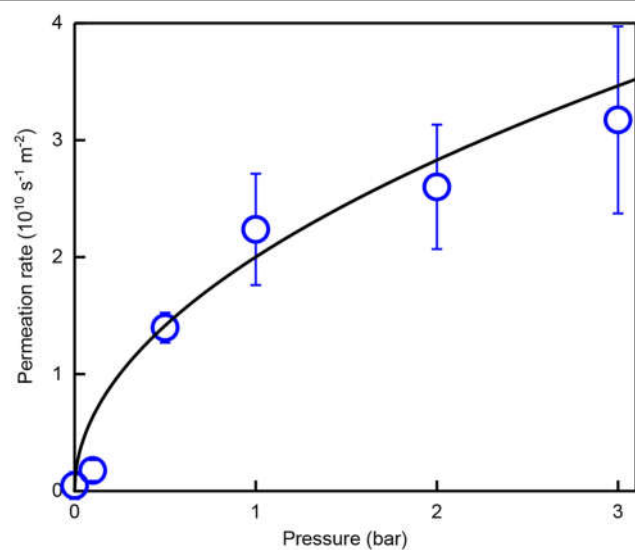
a, Changes in $\delta(0)$ measured for two containers with $d = 1 \mu\text{m}$; AFM scans were taken every hour. **b**, Representative profiles $\delta(x)$ for **a**. **c**, Long-term variations in δ for 12 different containers kept in air. **d**, Permeation rates evaluated from the evolution of $\Delta\delta$ with time in **c**. **e**, $\Delta\delta$ for 16 different devices placed in helium at 3 bar. **f**, Permeation rates for the data in **e**. In **c**, **e**, different symbols denote

different microcontainers made from graphite (empty symbols) and hBN (solid). The dashed lines in **c**, **e** indicate maximum changes detected for representative devices (colour coded). In **d**, **f**, the colour represents the same-colour device as in **c**, **e**, respectively. Error bars are standard deviation for fitting δ with a linear time dependence. Grey areas are the overall statistical accuracy obtained using all our devices measured in air and helium at 3 bar.



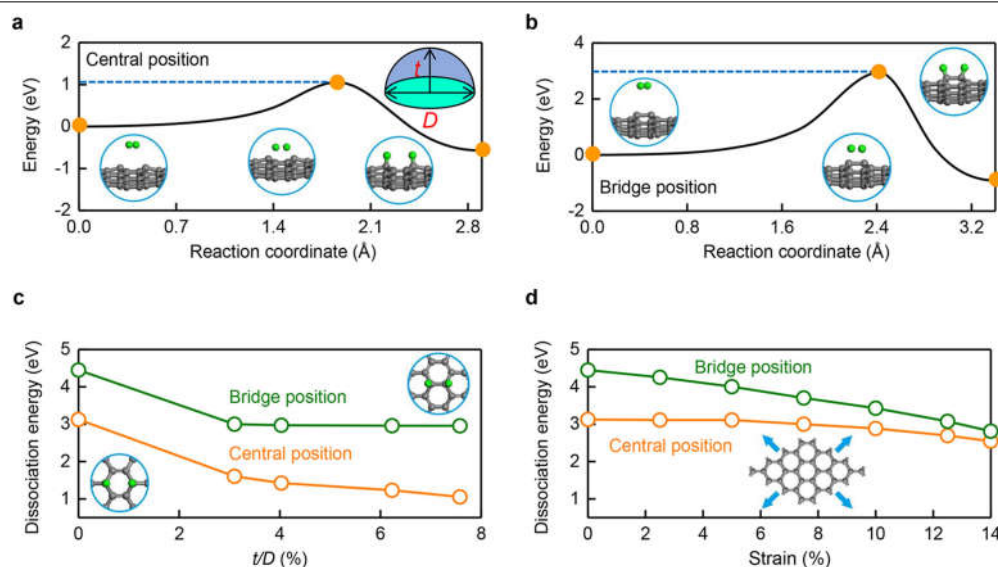
Extended Data Fig. 4 | Bilayer graphene and monolayer MoS_2 are impermeable to hydrogen. **a**, AFM micrographs of the same container sealed with bilayer graphene before (left) and after (right) its exposure to molecular hydrogen at 1 bar at 50 °C for 3 d. White curves show the profiles along the membrane's diameter. No changes in membrane positions could be detected within our experimental accuracy. **b**, Same experiment for monolayer MoS_2 . No changes could be noticed either. **c**, For comparison, we show the simultaneous

experiment for a microcontainer covered with monolayer graphene. The membrane clearly bulged out after the exposure, similar to the case of Fig. 2a. All scale bars, 1 μm . After the experiment, the bulging membrane in **c** was kept under ambient conditions and found to slowly deflate over months, in agreement with the room-temperature permeation rates reported in the main text.



Extended Data Fig. 5 | Pressure dependence of hydrogen permeation.

Symbols are measurements at room temperature. Error bars are standard deviation using a minimum of ten devices in each case. Solid curve is the best fit to the square-root dependence.



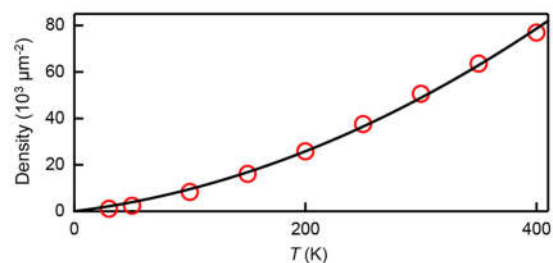
Extended Data Fig. 6 | Dissociation of molecular hydrogen at graphene ripples. **a, b**, Reaction of molecular hydrogen with graphene for $t/D = 7.5\%$ if adatoms are adsorbed in the central (**a**) and bridge (**b**) positions. Insets: atomic configurations of the initial, maximum-energy and final states (marked by the orange dots). The simulated size D is 2–3 nm. **c**, The dissociation barrier as a

function of ripples' curvature. Insets show top view of the bridge and central positions for hydrogen adatoms. **d**, The dissociation energy barrier as a function of biaxial strain. The inset shows schematic showing the direction of applied strain in our simulations.

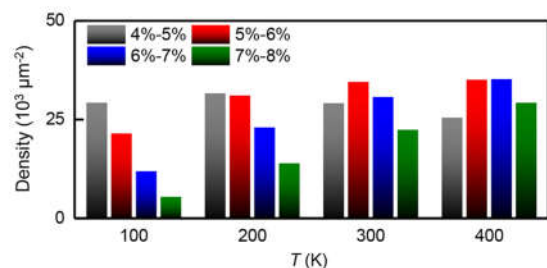
a



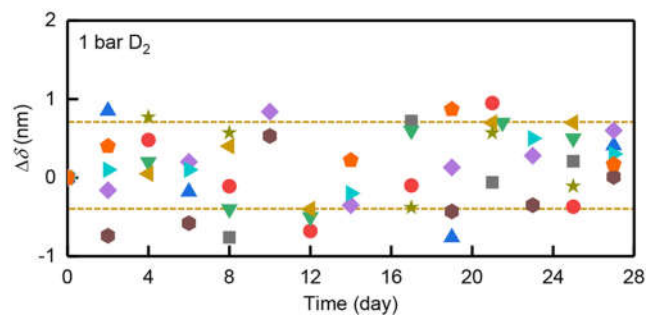
b



c



Extended Data Fig. 7 | Intrinsic (dynamic) ripples in graphene at different temperatures. **a**, Typical snapshot of graphene membrane at 300 K using molecular dynamics simulations. **b**, Density of ripples with $t/D \geq 7\%$ (most chemically active). Symbols are the calculations for different T . Solid curve is a guide to the eye. **c**, Statistical distribution of intrinsic ripples with different t/D .



Extended Data Fig. 8 | Isotope effect. Time evolution of $\Delta\delta$ for ten different devices (different colours) exposed to deuterium at 1 bar at room temperature (295 ± 2 K). All the devices are hBN containers sealed with monolayer graphene. The dashed lines indicate maximum changes detected for the device coded with the same colour. The random fluctuations are close in amplitude to those shown in Fig. 1e for helium.

Mass balance of the Greenland Ice Sheet from 1992 to 2018

<https://doi.org/10.1038/s41586-019-1855-2>

The IMBIE Team*

Received: 15 August 2019

Accepted: 25 November 2019

Published online: 10 December 2019

 Check for updates

The Greenland Ice Sheet has been a major contributor to global sea-level rise in recent decades^{1,2}, and it is expected to continue to be so³. Although increases in glacier flow^{4–6} and surface melting^{7–9} have been driven by oceanic^{10–12} and atmospheric^{13,14} warming, the magnitude and trajectory of the ice sheet's mass imbalance remain uncertain. Here we compare and combine 26 individual satellite measurements of changes in the ice sheet's volume, flow and gravitational potential to produce a reconciled estimate of its mass balance. The ice sheet was close to a state of balance in the 1990s, but annual losses have risen since then, peaking at 345 ± 66 billion tonnes per year in 2011. In all, Greenland lost $3,902 \pm 342$ billion tonnes of ice between 1992 and 2018, causing the mean sea level to rise by 10.8 ± 0.9 millimetres. Using three regional climate models, we show that the reduced surface mass balance has driven $1,964 \pm 565$ billion tonnes (50.3 per cent) of the ice loss owing to increased meltwater runoff. The remaining $1,938 \pm 541$ billion tonnes (49.7 per cent) of ice loss was due to increased glacier dynamical imbalance, which rose from 46 ± 37 billion tonnes per year in the 1990s to 87 ± 25 billion tonnes per year since then. The total rate of ice loss slowed to 222 ± 30 billion tonnes per year between 2013 and 2017, on average, as atmospheric circulation favoured cooler conditions¹⁵ and ocean temperatures fell at the terminus of Jakobshavn Isbræ¹⁶. Cumulative ice losses from Greenland as a whole have been close to the rates predicted by the Intergovernmental Panel on Climate Change for their high-end climate warming scenario¹⁷, which forecast an additional 70 to 130 millimetres of global sea-level rise by 2100 compared with their central estimate.

The Greenland Ice Sheet holds enough water to raise mean global sea level by 7.4 m (ref. ¹⁸). Its ice flows to the oceans through a network of glaciers and ice streams¹⁹, each with a substantial inland catchment²⁰. Fluctuations in the mass of the Greenland Ice Sheet occur due to variations in snow accumulation, meltwater runoff, ocean-driven melting and iceberg calving. There have been marked increases in air²¹ and ocean¹² temperatures and reductions in summer cloud cover²² around Greenland in the past few decades. These changes have produced increases in surface runoff⁸, supraglacial lake formation²³ and drainage²⁴, iceberg calving²⁵, glacier terminus retreat²⁶, submarine melting^{10,11} and ice flow⁶, leading to widespread changes in the surface elevation—particularly near the margin of the ice sheet (Fig. 1).

Over recent decades, ice losses from Greenland have made a substantial contribution to global sea-level rise², and model projections suggest that this imbalance will continue in a warming climate³. Since the early 1990s there have been comprehensive satellite observations of changing ice sheet velocity^{4,6}, elevation^{27–29} and, between 2002 and 2016, its changing gravitational attraction^{30,31}, from which complete estimates of Greenland Ice Sheet mass balance are determined¹. Before the 1990s, only partial surveys of the ice sheet elevation³² and velocity³³ change are available. In combination with models of surface mass balance (SMB; the net difference between precipitation, sublimation and meltwater runoff) and glacial isostatic adjustment³⁴, satellite

measurements, reported by the 2012 Ice Sheet Mass Balance Inter-comparison Exercise (IMBIE)¹, have shown a fivefold increase in the rate of ice loss from Greenland overall, rising from 51 ± 65 Gt yr⁻¹ in the early 1990s to 263 ± 30 Gt yr⁻¹ between 2005 and 2010. This ice loss has been driven by changes in SMB^{7,21} and ice dynamics^{5,33}. There was, however, a marked reduction in ice loss between 2013 and 2018, as a consequence of cooler atmospheric conditions and increased precipitation¹⁵. Although the broad pattern of change across Greenland (Fig. 1) is one of ice loss, there is considerable variability; for example, during the 2000s just four glaciers were responsible for half of the total ice loss due to increased discharge⁵, whereas many others contribute today³³. Moreover, some neighbouring ice streams have been observed to speed up over this period while others slowed down³⁵, suggesting diverse reasons for the changes that have taken place—including their geometrical configuration and basal conditions, as well as the forcing they have experienced³⁶. In this study we combine satellite altimetry, gravimetry and ice velocity measurements to produce a reconciled estimate of the Greenland Ice Sheet mass balance between 1992 and 2018, we evaluate the impact of changes in SMB and uncertainty in glacial isostatic adjustment and we partition the ice sheet mass loss into signals associated with surface mass balance and ice dynamics. In doing so, we extend a previous assessment¹ to include more satellite and ancillary data and to cover the period since 2012.

*A list of participants and their affiliations appears at the end of the paper

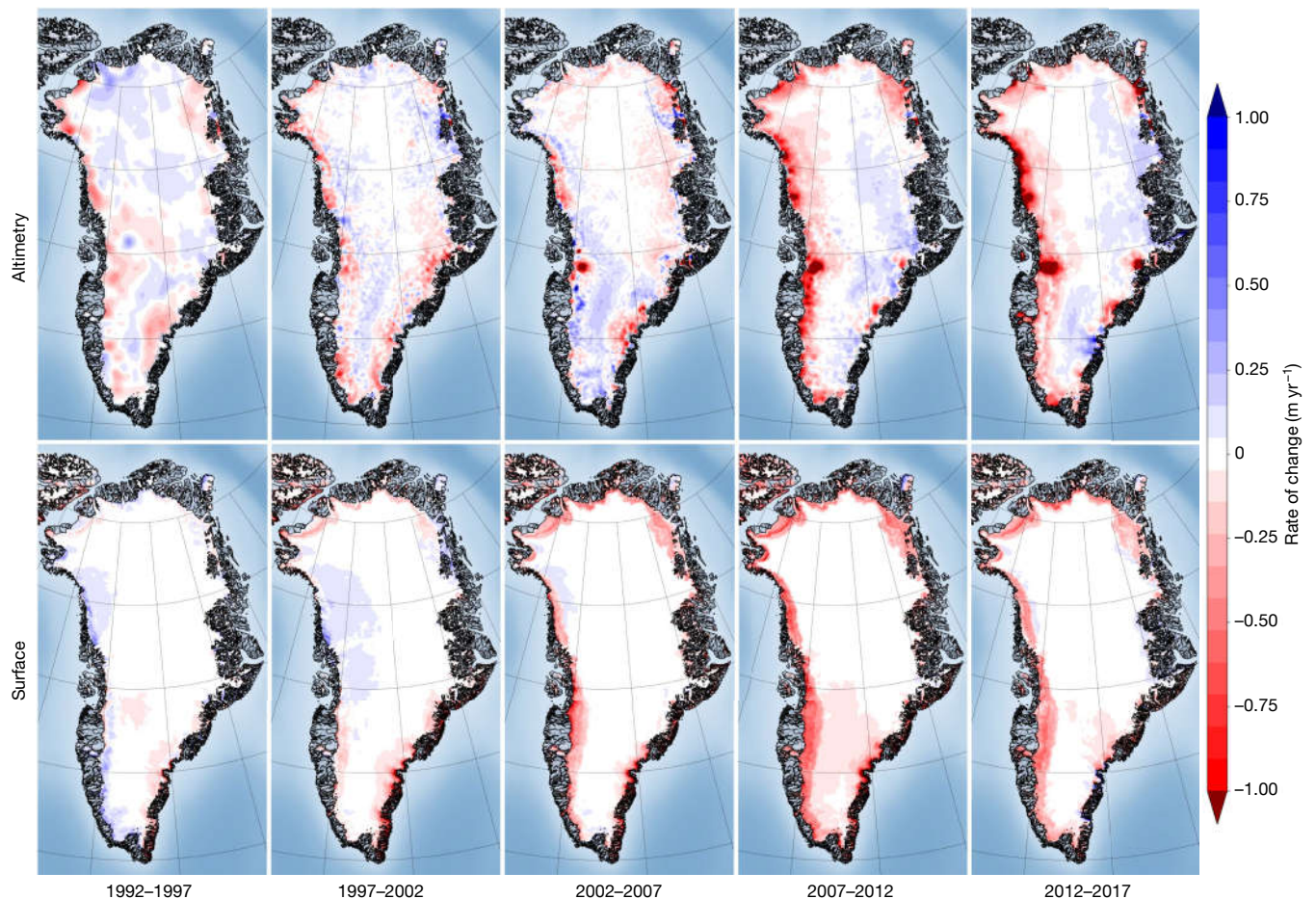


Fig. 1 | Greenland Ice Sheet elevation change. Rate of elevation change of the Greenland Ice Sheet determined from ERS, ENVISAT and CryoSat-2 satellite radar altimetry (top row) and from the HIRHAM5 SMB model (ice equivalent; bottom row) over successive 5-yr epochs. Data from ref. ²⁹.

Data and methods

We use 26 estimates of ice sheet mass balance derived from satellite altimetry (9 datasets), satellite gravimetry (14 datasets) and the input–output method (3 datasets) to assess changes in the Greenland Ice Sheet mass balance. The satellite data were computed using common spatial^{20,37} and temporal domains, and a range of models to estimate signals associated with changes in SMB and glacial isostatic adjustment. Satellite altimetry provides direct measurements of changing ice sheet surface elevation recorded at orbit crossing points³², along repeated ground tracks²⁷ or using plane-fit solutions²⁸. The ice sheet mass balance is estimated from these measurements either by prescribing the density of the elevation fluctuation³⁸ or by making an explicit model-based correction for changes in firn height³⁹. Satellite gravimetry measures fluctuations in the Earth's gravitational field computed using either global spherical harmonic solutions³⁰ or using spatially discrete mass concentration units³¹. Ice sheet mass changes are determined after making model-based corrections for glacial isostatic adjustment³⁰. The input–output method uses model estimates of SMB⁷ (the input) and satellite observations of ice sheet velocity computed from radar⁶ and optical⁴⁰ imagery combined with airborne measurements of ice thickness³³ to compute changes in marine-terminating glacier discharge into the oceans (the output). The overall mass balance is the difference between the input and output. Not all annual surveys of ice sheet discharge are complete, and sometimes regional extrapolations have to be employed to account for gaps in coverage³³. Because they provide important ancillary data, we also assess six models of glacial isostatic adjustment and ten models of surface mass balance.

To compare and aggregate the individual satellite datasets, we first adopt a common approach to derive linear rates of ice sheet mass balance over 36-month intervals (see Methods). We then compute error-weighted averages of all altimetry, gravimetry and input–output group mass trends, and combine these into a single reconciled estimate of the ice sheet mass balance using error-weighting of the group trends. Uncertainties in the individual rates of mass change are estimated as the root sum square of the linear model misfit and their measurement error, uncertainties in the group rates are estimated as the root mean square of the contributing time-series errors and uncertainties in the reconciled rates are estimated as their root mean square error divided by the square root of the number of independent groups. Cumulative uncertainties are computed as the root sum square of annual errors, an approach that has been employed in numerous studies^{1,17,33,41} and assumes that annual errors are not correlated over time. To improve on this assumption, it is necessary to consider the covariance of the systematic and random errors present in each mass balance solution (see Methods).

Intercomparison of satellite and model results

The satellite gravimetry and satellite altimetry data used in our assessment are corrected for the effects of glacial isostatic adjustment, although the correction is relatively small for altimetry as it manifests as a change in elevation and not mass. The most prominent and consistent local signals of glacial isostatic adjustment among the six models we considered are two instances of uplift peaking at about 5–6 mm yr⁻¹, one centred over northwest Greenland and Ellesmere Island, and one

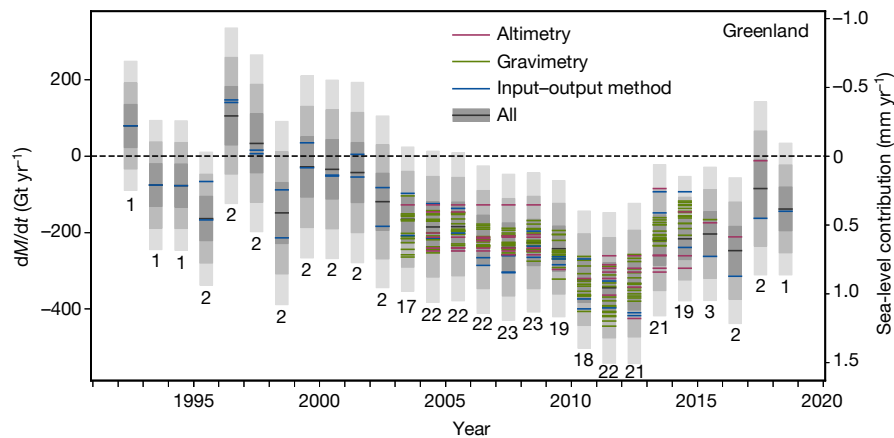


Fig. 2 | Greenland Ice Sheet mass balance. Rate of mass change (dM/dt , where M is mass and t time) of the Greenland Ice Sheet determined from the satellite altimetry, input–output method and gravimetry assessments included in this study. In each case, dM/dt is computed at annual intervals from time series of relative mass change using a 3-yr window. An average of the estimates across each measurement technique is also shown for each year (black line). The

estimated 1σ , 2σ and 3σ ranges of the class average are shaded in dark, mid and light grey, respectively; 97% of all estimates fall within the 1σ range, given their estimated individual errors. The equivalent sea-level contribution of the mass change is also indicated (right vertical axis), and the number of individual mass-balance estimates collated at each epoch is shown below each bar.

over northeast Greenland (see Methods and Extended Data Fig. 3). Although some models identify a 2 mm yr^{-1} subsidence under large parts of the central and southern parts of the ice sheet, it is absent or of lower magnitude in others, which suggests that it is less certain (Extended Data Table 1). The greatest difference among model solutions is at Kangerlussuaq Glacier in the southeast, where a study⁴² has shown that models and observations agree if a localized weak Earth structure associated with overpassing the Iceland hotspot is assumed; the effect is to offset earlier estimates of mass trends associated with glacial isostatic adjustment by about 20 Gt yr^{-1} . Farther afield, the highest spread between modelled uplift occurs on Baffin Island and beyond due to variations in regional model predictions related to the demise of the Laurentide Ice Sheet⁴². This regional uncertainty is probably a major factor in the spread across the ice-sheet-wide estimates. Nevertheless, at $-3 \pm 20 \text{ Gt yr}^{-1}$, the mass signal associated with glacial isostatic adjustment in Greenland shows no coherent substantive change and is negligible relative to reported ice sheet mass trends¹.

There is generally good agreement between the models of Greenland Ice Sheet SMB that we have assessed for determining mass input–particularly those of a similar class; for example, 70% of all model estimates of runoff and accumulation fall within 1σ of their mean (see Methods and Extended Data Table 2). The exceptions are a global reanalysis with coarse spatial resolution that tends to underestimate runoff due to its poor delineation of the ablation zone, and a snow process model that tends to underestimate precipitation and to overestimate runoff in most sectors. Among the other eight models, the average surface mass balance between 1980 and 2012 is $361 \pm 40 \text{ Gt yr}^{-1}$, with a marked negative trend over time (Extended Data Fig. 4) that is mainly due to increased runoff⁷. At the regional scale, the largest differences occur in the northeast, where two regional climate models predict considerably less runoff, and in the southeast, where there is considerable spread in precipitation and runoff across all models. All models show high temporal variability in SMB components, and all models show that the southeast receives the highest net intake of mass at the surface due to high rates of snowfall originating from the Icelandic Low⁴³. By contrast, the southwest, which features the widest ablation zone⁷, has experienced alternate periods of net surface mass loss and gain over recent decades, and has the lowest average SMB across the ice sheet.

We assessed the consistency of the satellite altimetry, gravimetry and input–output method estimates of Greenland Ice Sheet mass balance using common spatial and temporal domains (see Fig. 2 and Methods).

In general, there is close agreement between estimates determined using each approach, and the standard deviations of annual mass balance solutions from the coincident altimetry, gravimetry and input–output methods are 42 , 31 and 23 Gt yr^{-1} , respectively (Extended Data Table 3). Once averages were computed for each technique, the resulting estimates of mass balance were also closely aligned (Extended Data Fig. 6). For example, over the common period 2005–2015, the average Greenland Ice Sheet mass balance is $-254 \pm 18 \text{ Gt yr}^{-1}$ and, by comparison, the spread of the altimetry, gravimetry and input–output method estimates is just 36 Gt yr^{-1} (Extended Data Table 3). The estimated uncertainty of the aggregated mass balance solution (see Methods) is larger than the standard deviation of model corrections for glacial isostatic adjustment (20 Gt yr^{-1} for gravimetry) and for surface mass balance (40 Gt yr^{-1}), which suggests that their collective impacts have been adequately compensated; it is also larger than the estimated 30 Gt yr^{-1} mass losses from peripheral ice caps⁴⁴, which are not accounted for in all individual solutions. In keeping with results from Antarctica⁴¹, rates of mass loss determined using the input–output method are the most negative, and those determined from altimetry are the least negative. However, the spread among the three techniques is six times lower for Greenland than it is for Antarctica⁴¹, reflecting differences in ice sheet size, the complexity of the mass balance processes and the limitations of the various geodetic techniques.

Ice sheet mass balance

We aggregated the average mass balance estimates from gravimetry, altimetry and the input–output method to form a single, time-varying record (Fig. 2) and then integrated these data to determine the cumulative mass lost from Greenland since 1992 (Fig. 3). Although Greenland has been losing ice throughout most of the intervening period, the rate of loss has varied considerably. The rate of ice loss progressively increased between 1992 and 2012, reaching a maximum of $345 \pm 66 \text{ Gt yr}^{-1}$ in 2011, ahead of the extreme summertime surface melting that occurred in the following year¹⁴. Since 2012, however, the trend has reversed, with a progressive reduction in the rate of mass loss during the subsequent period. By 2018—the last complete year of our survey—the annual rate of ice mass loss had reduced to $85 \pm 75 \text{ Gt yr}^{-1}$. The highly variable nature of ice losses from Greenland is a consequence of the wide range of physical processes that are affecting different sectors of the ice sheet^{16,28,35}, which suggests that care should be taken when extrapolating measurements that

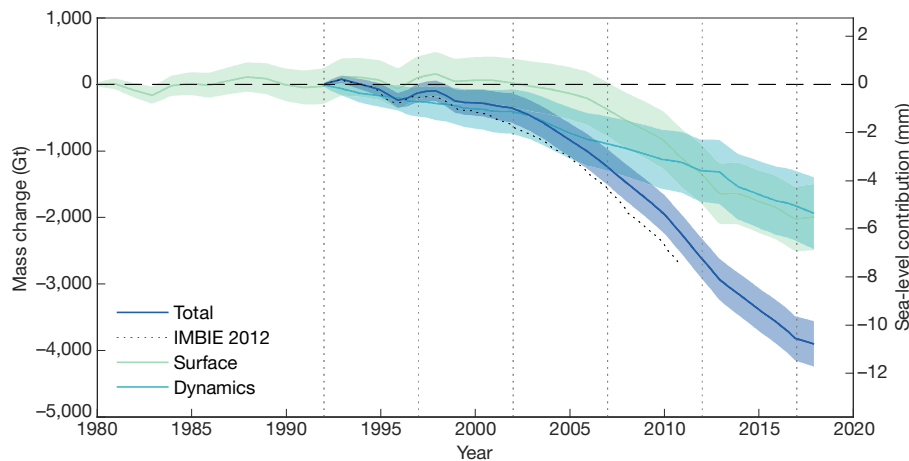


Fig. 3 | Cumulative anomalies in the total mass, SMB and ice dynamics of the Greenland Ice Sheet. The total change is determined as the integral of the average rate of ice sheet mass change (Fig. 2). The change in SMB is determined from three regional climate models relative to their mean over the period 1980–1990. The change associated with ice dynamics is determined as the difference between the change in total and surface mass. The estimated 1σ

uncertainties of the cumulative changes are shown by the shaded envelopes. The dotted line shows the result of a previous assessment¹. The equivalent sea-level contribution of the mass change is also indicated (right vertical axis). Vertical dashed lines mark consecutive 5-yr epochs since the start of our satellite record in 1992. The IMBIE 2012 data is from ref. ¹.

are sparse in space or time. Although the rates of mass loss we have computed between 1992 and 2011 are 16% less negative than those of a previous assessment, which included far fewer datasets¹, the results are consistent given their respective uncertainties. Altogether, the Greenland Ice Sheet has lost $3,902 \pm 342$ Gt of ice to the ocean since 1992, with roughly half of this loss occurring during the 6-yr period between 2006 and 2012.

To determine the proportion of mass lost due to surface and ice dynamical processes, we computed the contemporaneous trend in Greenland Ice Sheet surface mass balance—the net balance between precipitation and ablation⁷, which is controlled by interactions with the atmosphere (Fig. 3). In Greenland, recent trends in surface mass balance have been largely driven by meltwater runoff¹³, which has increased as the regional climate has warmed¹³. As direct observations of ice sheet surface mass balance are too scarce to provide full temporal and spatial coverage⁴⁵, regional estimates are usually taken from atmospheric models that are evaluated with existing observations. Our evaluation (see Methods) shows that the finer-spatial-resolution regional climate models produce consistent results, probably due to their ability to capture local changes in melting and precipitation associated with atmospheric forcing, and to resolve the full extent of the ablation zone⁴⁶. We therefore compare and combine estimates of Greenland SMB derived from three regional climate models: RACMO2.3p2⁴⁶, MARv3.6²¹ and HIRHAM⁹. To assess the surface mass change across the Greenland Ice Sheet between 1980 and 2018, we accumulate SMB anomalies from each of the regional climate models (Extended Data Fig. 7) and average them into a single estimate (Fig. 3). These SMB anomalies are computed with respect to the average between 1980 and 1990, which corresponds to a period of approximate balance⁸ and is

common to all models. In this comparison, all three models show that the Greenland Ice Sheet entered abruptly into a period of anomalously low SMB in the late 1990s and, when combined, they show that the ice sheet lost $1,964 \pm 565$ Gt of its mass due to meteorological processes between 1992 and 2018 (Table 1).

Just over half (50.3%) of all mass losses from Greenland—and much of their short-term variability—have been due to variations in the ice sheet’s SMB and its indirect impacts on firn processes. For example, between 2007 and 2012, 70% of the total ice loss (193 ± 37 Gt yr^{−1}) was due to SMB, compared with 27% (22 ± 20 Gt yr^{−1}) over the preceding 15 years and 57% (139 ± 38 Gt yr^{−1}) since then (Table 1). The rise in the total rate of ice loss during the late 2000s coincided with warmer atmospheric conditions, which promoted several episodes of widespread melting and runoff¹⁴. The reduction in surface mass loss since then is associated with a shift of the North Atlantic Oscillation, which brought about cooler atmospheric conditions and increased precipitation along the southeastern coast¹⁵. Trends in the total ice sheet mass balance are not entirely due to surface mass balance, however, and by calculating the difference between these two signals, we can estimate the total change in mass loss due to ice dynamical imbalance—that is, the integrated net mass loss from those glaciers whose velocity does not equal their long-term mean (Fig. 3). Although this approach is indirect, it makes use of all the satellite observations and regional climate models included in our study, overcoming limitations in the spatial and temporal sampling of ice discharge estimates derived from ice velocity and thickness data. Our estimate shows that, between 1992 and 2018, Greenland lost $1,938 \pm 541$ Gt of ice due to the dynamical imbalance of glaciers relative to their steady state, accounting for 49.7% of the total imbalance (Table 1). Losses due to increased ice discharge rose

Table 1 | Rates of Greenland Ice Sheet total, surface and dynamical mass change

Region	1992–1997 (Gt yr ^{−1})	1997–2002 (Gt yr ^{−1})	2002–2007 (Gt yr ^{−1})	2007–2012 (Gt yr ^{−1})	2012–2017 (Gt yr ^{−1})	1992–2011 (Gt yr ^{−1})	1992–2018 (Gt yr ^{−1})
Total	-26 ± 27	-44 ± 35	-174 ± 30	-275 ± 28	-244 ± 28	-119 ± 16	-150 ± 13
Surface	26 ± 35	-15 ± 36	-78 ± 36	-193 ± 37	-139 ± 38	-57 ± 18	-76 ± 16
Dynamics	-52 ± 44	-29 ± 50	-96 ± 47	-82 ± 46	-105 ± 47	-62 ± 24	-75 ± 21

Total rates were determined from all satellite measurements over the various epochs, rates of surface mass change were determined from three regional climate models and rates of dynamical mass change were determined as the difference between the two. The period 1992–2011 is included for comparison with a previous assessment¹, which reported a mass balance estimate of -142 ± 49 Gt yr^{−1} based on far fewer data. The small differences in our updated estimate are due to our inclusion of more data and an updated aggregation scheme (see Methods). Errors are 1σ.

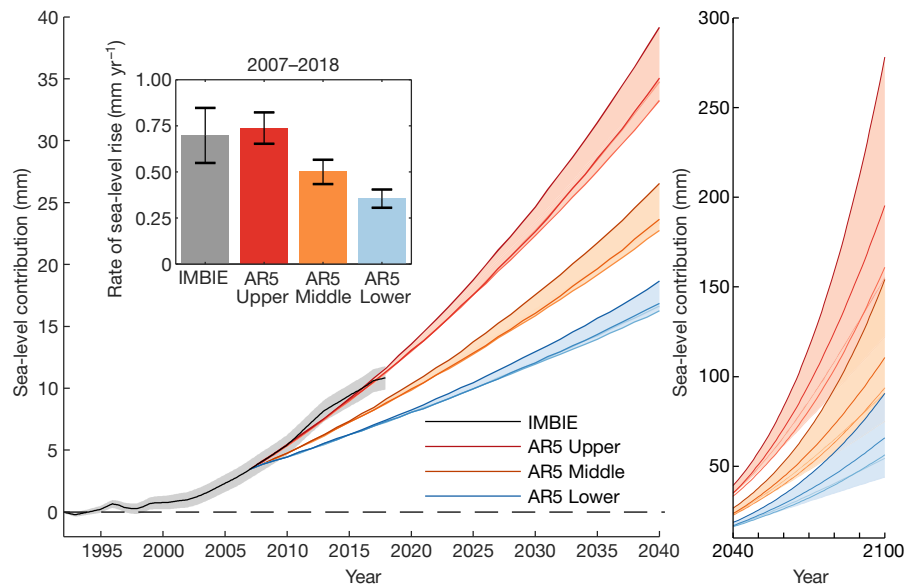


Fig. 4 | Observed and predicted sea-level contributions from Greenland Ice Sheet mass change. The global sea-level contribution from Greenland Ice Sheet mass change according to this study and the IPCC AR5 projections between 1992–2040 (left) and 2040–2100 (right) including upper, mid and lower estimates from the sum of modelled SMB and rapid ice dynamical contributions. Darker lines represent pathways from the five AR5 scenarios in

order of increasing emissions: RCP2.6, RCP4.5, RCP6.0, SRES A1B and RCP8.5. Shaded areas represent the spread of AR5 emissions scenarios and the 1 σ estimated error on the IMBIE data (this study). Inset, the average annual rates of sea-level rise during the overlap period 2007–2018 and their standard deviations (error bars). Cumulative AR5 projections have been offset to make them equal to the observational record at their start date (2007).

sharply in the early 2000s when Jakobshavn Isbræ¹⁰ and several other outlet glaciers in the southeast⁴⁷ sped up, and the discharge losses are now four times higher than in the 1990s. For the period between 2002 and 2007, ice dynamical imbalance was the major source of ice loss from the ice sheet as a whole, although the situation has since returned to being dominated by surface mass losses as several glaciers have slowed down¹⁶.

Despite a reduction in the overall rate of ice loss from Greenland between 2013 and 2018 (Fig. 2), the ice sheet mass balance remained negative, adding 10.8 ± 0.9 mm to global sea level since 1992. Although the average sea level contribution is 0.42 ± 0.04 mm yr⁻¹, the 5-yr average rate varied by a factor of 5 over the 25-yr period, peaking at 0.76 ± 0.08 mm yr⁻¹ between 2007 and 2012. The variability in ice loss from Greenland illustrates the importance of accounting for annual fluctuations when attempting to close the global sea-level budget². Satellite records of ice sheet mass balance are also an important tool for evaluating numerical models of ice sheet evolution⁴⁸. In their 2013 assessment, the Intergovernmental Panel on Climate Change (IPCC) predicted ice losses from Greenland due to SMB and glacier dynamics under a range of scenarios, beginning in 2007¹⁷ (Fig. 4). Although ice losses from Greenland have fluctuated considerably during the 12-yr period of overlap between the IPCC predictions and our reconciled time series, the total change and average rate (0.70 mm yr⁻¹) are close to the upper range of predictions (0.72 mm yr⁻¹), which implies 70–130 mm of sea-level rise by 2100 above central estimates. The drop in ice losses between 2013 and 2018, however, shifted rates towards the lower end of projections, and a longer period of comparison is required to establish whether the upper trajectory will continue to be followed. Even greater sea-level contributions cannot be ruled out if feedbacks between the ice sheet and other elements of the climate system are underestimated by current ice sheet models³. Although the volume of ice stored in Greenland is a small fraction of that in Antarctica (12%), its recent losses have been ~38% higher⁴¹ as a consequence of the relatively strong atmospheric^{13,14} and oceanic^{10,11} warming that has occurred in its vicinity, and it is expected to continue to be a major source of sea-level rise^{3,17}.

Conclusions

We combine 26 satellite estimates of ice sheet mass balance and assess 10 models of ice sheet SMB and 6 models of glacial isostatic adjustment to show that the Greenland Ice Sheet lost $3,902 \pm 342$ Gt of ice between 1992 and 2018. During the common period 2005–2015, the spread of mass balance estimates derived from three techniques is 36 Gt yr⁻¹, or 14% of the estimated rate of imbalance. The rate of ice loss has generally increased over time, rising from 26 ± 27 Gt yr⁻¹ between 1992 and 1997, peaking at 275 ± 28 Gt yr⁻¹ between 2007 and 2012, and reducing to 244 ± 28 Gt yr⁻¹ between 2012 and 2017. Just over half ($1,964 \pm 565$ Gt, or 50.3%) of the ice losses are due to reduced SMB (mostly meltwater runoff) associated with changing atmospheric conditions^{13,14}, and these changes have also driven the shorter-term temporal variability in ice sheet mass balance. Despite variations in the imbalance of individual glaciers^{4,5,33}, ice losses due to increasing discharge from the ice sheet as a whole have risen steadily from 46 ± 37 Gt yr⁻¹ in the 1990s to 87 ± 25 Gt yr⁻¹ since then, and account for just under half of all losses (49.7%) over the survey period.

Our assessment shows that estimates of Greenland Ice Sheet mass balance derived from satellite altimetry, gravimetry and the input–output method agree to within 20 Gt yr⁻¹, that model estimates of SMB agree to within 40 Gt yr⁻¹ and that model estimates of glacial isostatic adjustment agree to within 20 Gt yr⁻¹. These differences represent a small fraction (13%) of the Greenland Ice Sheet mass imbalance and are comparable to its estimated uncertainty (13 Gt yr⁻¹). Nevertheless, there is still disagreement among models of glacial isostatic adjustment in northern Greenland. Spatial resolution is a key factor in the degree to which models of SMB can represent ablation and precipitation at local scales, and estimates of ice sheet mass balance determined from satellite altimetry and the input–output method continue to be positively and negatively biased, respectively, compared with those based on satellite gravimetry (albeit by small amounts). More satellite estimates of ice sheet mass balance at the start (1990s) and end (2010s) of our record would help to reduce the dependence on fewer data during those periods; although new missions^{49,50} will no doubt address the latter period, further analysis of historical satellite data are required to address the paucity of data during the early years.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1855-2>.

- Shepherd, A. et al. A reconciled estimate of ice-sheet mass balance. *Science* **338**, 1183–1189 (2012).
- WCRP Global Sea Level Budget Group. Global sea-level budget 1993–present. *Earth Syst. Sci. Data* **10**, 1551–1590 (2018).
- Pattyn, F. et al. The Greenland and Antarctic ice sheets under 1.5 °C global warming. *Nat. Clim. Change* **8**, 1053–1061 (2018).
- Moon, T., Joughin, I., Smith, B. & Howat, I. 21st-century evolution of Greenland outlet glacier velocities. *Science* **336**, 576–578 (2012).
- Enderlin, E. M. et al. An improved mass budget for the Greenland ice sheet. *Geophys. Res. Lett.* **41**, 866–872 (2014).
- Rignot, E. & Kanagaratnam, P. Changes in the velocity structure of the Greenland Ice Sheet. *Science* **311**, 986–990 (2006).
- van den Broeke, M. et al. Partitioning recent Greenland mass loss. *Science* **326**, 984–986 (2009).
- Trusel, L. D. et al. Nonlinear rise in Greenland runoff in response to post-industrial Arctic warming. *Nature* **564**, 104–108 (2018).
- Lucas-Picher, P. et al. Very high resolution regional climate model simulations over Greenland: identifying added value. *J. Geophys. Res.* **D 117**, 02108 (2012).
- Holland, D. M., Thomas, R. H., de Young, B., Ribergaard, M. H. & Lyberth, B. Acceleration of Jakobshavn Isbræ triggered by warm subsurface ocean waters. *Nat. Geosci.* **1**, 659–664 (2008).
- Seale, A., Christoffersen, P., Mugford, R. I. & O’Leary, M. Ocean forcing of the Greenland Ice Sheet: calving fronts and patterns of retreat identified by automatic satellite monitoring of eastern outlet glaciers. *J. Geophys. Res. Earth Surf.* **116**, F03013 (2011).
- Straneo, F. & Heimbach, P. North Atlantic warming and the retreat of Greenland’s outlet glaciers. *Nature* **504**, 36–43 (2013).
- Hanna, E., Mernild, S. H., Cappelen, J. & Steffen, K. Recent warming in Greenland in a long-term instrumental (1881–2012) climatic context: I. Evaluation of surface air temperature records. *Environ. Res. Lett.* **7**, 045404 (2012).
- Fettweis, X. et al. Important role of the mid-tropospheric atmospheric circulation in the recent surface melt increase over the Greenland ice sheet. *Cryosphere* **7**, 241–248 (2013).
- Bevis, M. et al. Accelerating changes in ice mass within Greenland, and the ice sheet’s sensitivity to atmospheric forcing. *Proc. Natl Acad. Sci. USA* **116**, 1934–1939 (2019).
- Khazendar, A. et al. Interruption of two decades of Jakobshavn Isbræ acceleration and thinning as regional ocean cools. *Nat. Geosci.* **12**, 277–283 (2019); correction **12**, 493 (2019).
- Church, J. A. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 1137–1216 (IPCC, Cambridge Univ. Press, 2013).
- Morlighem, M. et al. BedMachine v3: complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation. *Geophys. Res. Lett.* **44**, 11,051–11,061 (2017).
- Joughin, I., Smith, B. E., Howat, I. M., Scambos, T. & Moon, T. Greenland flow variability from ice-sheet-wide velocity mapping. *J. Glaciol.* **56**, 415–430 (2010).
- Zwally, H. J., Giovinetto, M. B., Beckley, M. A. & Saba, J. L. *Antarctic and Greenland Drainage Systems* (GSFC Cryospheric Sciences Laboratory, 2012); http://icesat4.gsfc.nasa.gov/cryo_data/ant_grn_drainage_systems.php.
- Fettweis, X. et al. Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model. *Cryosphere* **11**, 1015–1033 (2017).
- Hofer, S., Tedstone, A. J., Fettweis, X. & Bamber, J. L. Decreasing cloud cover drives the recent mass loss on the Greenland Ice Sheet. *Sci. Adv.* **3**, e1700584 (2017).
- Leeson, A. A. et al. Supraglacial lakes on the Greenland ice sheet advance inland under warming climate. *Nat. Clim. Change* **5**, 51–55 (2015).
- Palmer, S., McMillan, M. & Morlighem, M. Subglacial lake drainage detected beneath the Greenland ice sheet. *Nat. Commun.* **6**, 8408 (2015).
- Nick, F. M. et al. The response of Petermann Glacier, Greenland, to large calving events, and its future stability in the context of atmospheric and oceanic warming. *J. Glaciol.* **58**, 229–239 (2012).
- Joughin, I. et al. Ice-front variation and tidewater behavior on Helheim and Kangerdlugssuaq Glaciers, Greenland. *J. Geophys. Res. Earth Surf.* **113**, F01004 (2008).
- Pritchard, H. D., Arthern, R. J., Vaughan, D. G. & Edwards, L. A. Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. *Nature* **461**, 971–975 (2009).
- McMillan, M. et al. A high-resolution record of Greenland mass balance. *Geophys. Res. Lett.* **43**, 7002–7010 (2016).
- Sandberg Sørensen, L. et al. 25 years of elevation changes of the Greenland Ice Sheet from ERS, Envisat, and CryoSat-2 radar altimetry. *Earth Planet. Sci. Lett.* **495**, 234–241 (2018).
- Velicogna, I. & Wahr, J. Greenland mass balance from GRACE. *Geophys. Res. Lett.* **32**, L18505 (2005).
- Luthcke, S. B. et al. Recent Greenland ice mass loss by drainage system from satellite gravity observations. *Science* **314**, 1286–1289 (2006).
- Zwally, H. J., Bindshadler, R. A., Brenner, A. C., Major, J. A. & Marsh, J. G. Growth of Greenland Ice Sheet: measurement. *Science* **246**, 1587–1589 (1989).
- Mouginot, J. et al. Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. *Proc. Natl Acad. Sci. USA* **116**, 9239–9244 (2019).
- Lecavalier, B. S. et al. A model of Greenland ice sheet deglaciation constrained by observations of relative sea level and ice extent. *Quat. Sci. Rev.* **102**, 54–84 (2014).
- King, M. D. et al. Seasonal to decadal variability in ice discharge from the Greenland Ice Sheet. *Cryosphere* **12**, 3813–3825 (2018).
- Porter, D. F. et al. Identifying spatial variability in Greenland’s outlet glacier response to ocean heat. *Front. Earth Sci.* **6**, 90 (2018).
- Rignot, E. & Mouginot, J. Ice flow in Greenland for the International Polar Year 2008–2009. *Geophys. Res. Lett.* **39**, L11501 (2012).
- Sørensen, L. S. et al. Mass balance of the Greenland ice sheet (2003–2008) from ICESat data—the impact of interpolation, sampling and firn density. *Cryosphere* **5**, 173–186 (2011).
- Zwally, H. J. et al. Greenland ice sheet mass balance: distribution of increased mass loss with climate warming; 2003–07 versus 1992–2002. *J. Glaciol.* **57**, 88–102 (2011).
- Rosenau, R., Scheinert, M. & Dietrich, R. A processing system to monitor Greenland outlet glacier velocity variations at decadal and seasonal time scales utilizing the Landsat imagery. *Remote Sens. Environ.* **169**, 1–19 (2015).
- The IMBIE Team. Mass balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature* **558**, 219–222 (2018).
- Khan, S. A. et al. Geodetic measurements reveal similarities between post-Last Glacial Maximum and present-day mass loss from the Greenland ice sheet. *Sci. Adv.* **2**, e1600931 (2016).
- Ettema, J. et al. Higher surface mass balance of the Greenland ice sheet revealed by high-resolution climate modeling. *Geophys. Res. Lett.* **36**, L12501 (2009).
- Bolch, T. et al. Mass loss of Greenland’s glaciers and ice caps 2003–2008 revealed from ICESat laser altimetry data. *Geophys. Res. Lett.* **40**, 875–881 (2013).
- Vernon, C. L. et al. Surface mass balance model intercomparison for the Greenland ice sheet. *Cryosphere* **7**, 599–614 (2013).
- Noël, B. et al. Modelling the climate and surface mass balance of polar ice sheets using RACMO2—Part 1: Greenland (1958–2016). *Cryosphere* **12**, 811–831 (2018).
- Howat, I. M., Joughin, I., Fahnestock, M., Smith, B. E. & Scambos, T. A. Synchronous retreat and acceleration of southeast Greenland outlet glaciers 2000–06: ice dynamics and coupling to climate. *J. Glaciol.* **54**, 646–660 (2008).
- Shepherd, A. & Nowicki, S. Improvements in ice-sheet sea-level projections. *Nat. Clim. Change* **7**, 672–674 (2017).
- Markus, T. et al. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation. *Remote Sens. Environ.* **190**, 260–273 (2017).
- Flechtner, F. et al. What can be expected from the GRACE-FO laser ranging interferometer for earth science applications? *Surv. Geophys.* **37**, 453–470 (2016).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

The IMBIE Team

Andrew Shepherd^{1,2,3}, Erik Ivins^{2,3}, Eric Rignot^{2,3}, Ben Smith⁴, Michiel van den Broeke⁵, Isabella Velicogna^{2,3}, Pippa Whitehouse⁶, Kate Briggs⁷, Ian Joughin⁸, Gerhard Krinner⁷, Sophie Nowicki⁹, Tony Payne⁹, Ted Scambos¹⁰, Nicole Schlegel², Geruo A³, Cécile Agosta¹¹, Andreas Ahlstrøm¹², Greg Babonis¹³, Valentina R. Barletta¹⁴, Anders A. Bjørk¹⁵, Alejandro Blazquez¹⁶, Jennifer Bonin¹⁷, William Colgan¹², Beata Csatho¹³, Richard Cullather¹⁸, Marcus E. Engdahl¹⁹, Denis Felikson⁸, Xavier Fettweis¹¹, Rene Forsberg¹⁴, Anna E. Hogg¹, Hubert Gallee², Alex Gardner², Lin Gilbert²⁰, Noel Gourmelen²¹, Andreas Groh²², Brian Gunter²³, Edward Hanna²⁴, Christopher Harig²⁵, Veit Helm²⁶, Alexander Horvath²⁷, Martin Horvath²², Shfaqat Khan¹⁴, Kristian K. Kjeldsen^{12,28}, Hannes Konrad²⁹, Peter L. Langen³⁰, Benoit Lecavalier³¹, Bryant Loomis⁸, Scott Luthcke⁸, Malcolm McMillan³², Daniele Melini³³, Sebastian Mernild^{34,35,36,37}, Yara Mohajerani³, Philip Moore³⁸, Ruth Mottram³⁰, Jeremie Mouginot^{3,7}, Gorka Moyano³⁹, Alan Muir²⁰, Thomas Nagler⁴⁰, Grace Nield⁶, Johan Nilsson⁴¹, Brice Noël⁵, Ines Otsaka¹, Mark E. Pattle³⁹, W. Richard Peltier⁴¹, Nadège Pie⁴², Roelof Rietbroek⁴³, Helmut Rott⁴⁰, Louise Sandberg Sørensen¹⁴, Ingo Sasgen²⁶, Himanshu Save⁴², Bernd Scheuchl³, Ernst Schrama⁴⁴, Ludwig Schröder^{22,26}, Ki-Weon Seo⁴⁵, Sebastian B. Simonsen¹⁴, Thomas Slater¹, Giorgio Spada⁴⁶, Tyler Sutterley³, Matthieu Talpe², Lev Tarasov³¹, Willem Jan van de Berg⁵, Wouter van der Wal^{44,47}, Melchior van Wessem⁵, Bramha Dutt Vishwakarma⁴⁸, David Wiese², David Wilton⁴⁹, Thomas Wagner⁵⁰, Bert Wouters^{5,47} & Jan Wuite⁴⁰

¹Centre for Polar Observation and Modelling, University of Leeds, Leeds, UK. ²NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. ³Department of Earth System Science, University of California, Irvine, CA, USA. ⁴Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA. ⁵Institute for Marine and Atmospheric Research, Utrecht University, Utrecht, The Netherlands. ⁶Department of Geography, Durham University, Durham, UK. ⁷Institute of Environmental Geosciences, Université Grenoble Alpes, Grenoble, France. ⁸Cryospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA. ⁹School of Geographical Sciences, University of Bristol, Bristol, UK. ¹⁰Earth Science and Observation Center, University of Colorado, Boulder, CO, USA. ¹¹Department of Geography, University of Liège, Liège, Belgium. ¹²Geological Survey of Denmark and Greenland, Copenhagen, Denmark. ¹³Department of Geology, State University of New York at Buffalo, Buffalo, NY, USA. ¹⁴DTU Space, National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark. ¹⁵Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark. ¹⁶LEGOS, Université de Toulouse, Toulouse, France. ¹⁷College of

Marine Sciences, University of South Florida, Tampa, FL, USA. ¹⁸Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA. ¹⁹ESA-ESRIN, Frascati, Italy. ²⁰Mullard Space Science Laboratory, University College London, Holmbury St Mary, UK. ²¹School of Geosciences, University of Edinburgh, Edinburgh, UK. ²²Institute for Planetary Geodesy, Technische Universität Dresden, Dresden, Germany. ²³Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ²⁴School of Geography, University of Lincoln, Lincoln, UK. ²⁵Department of Geosciences, University of Arizona, Tucson, AZ, USA. ²⁶Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. ²⁷Institute of Astronomical and Physical Geodesy, Technical University Munich, Munich, Germany. ²⁸GeoGenetics, Globe Institute, University of Copenhagen, Copenhagen, Denmark. ²⁹Deutscher Wetterdienst, Offenbach, Germany. ³⁰Danish Meteorological Institute, Copenhagen, Denmark. ³¹Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada. ³²Lancaster Environment Centre, University of Lancaster, Lancaster, UK. ³³Istituto Nazionale di Geofisica e

Vulcanologia, Rome, Italy. ³⁴Nansen Environmental and Remote Sensing Centre, Bergen, Norway. ³⁵Faculty of Engineering and Science, Western Norway University of Applied Sciences, Sogndal, Norway. ³⁶Direction of Antarctic and Sub-Antarctic Programs, Universidad de Magallanes, Punta Arenas, Chile. ³⁷Geophysical Institute, University of Bergen, Bergen, Norway. ³⁸School of Engineering, Newcastle University, Newcastle upon Tyne, UK. ³⁹IsardSAT, Barcelona, Catalonia. ⁴⁰ENVEO, Innsbruck, Austria. ⁴¹Department of Physics, University of Toronto, Toronto, Ontario, Canada. ⁴²Center for Space Research, University of Texas, Austin, TX, USA. ⁴³Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany. ⁴⁴Department of Space Engineering, Delft University of Technology, Delft, The Netherlands. ⁴⁵Department of Earth Science Education, Seoul National University, Seoul, South Korea. ⁴⁶Dipartimento di Scienze Pure e Applicate, Università di Urbino "Carlo Bo", Urbino, Italy. ⁴⁷Department of Civil Engineering, Delft University of Technology, Delft, The Netherlands. ⁴⁸Geodetic Institute, University of Stuttgart, Stuttgart, Germany. ⁴⁹Department of Computer Science, University of Sheffield, Sheffield, UK. ⁵⁰NASA Headquarters, Washington, DC, USA. ⁵¹e-mail: a.shepherd@leeds.ac.uk

Methods

Data

In this assessment, we analyse five groups of data: estimates of ice sheet mass-balance determined from three distinct classes of satellite observations (altimetry, gravimetry and the input–output method (IOM)) and model estimates of SMB) and glacial isostatic adjustment (GIA). Each dataset is computed following previously reported methods (based on refs. ^{28,33,38,51–119} and detailed in Supplementary Table 1) and, for consistency, they are aggregated within common spatial and temporal domains. Altogether, 26 separate ice sheet mass balance datasets were used—9 derived from satellite altimetry, 3 from the IOM and 14 from satellite gravimetry—with a combined period running from 1992 to 2018 (Extended Data Fig. 1). We also assess six model estimates of GIA (Extended Data Table 1) and ten model estimates of SMB (Extended Data Table 2).

Drainage basins

We analyse mass trends using two ice sheet drainage basin sets (Extended Data Fig. 2) for consistency with those used in the first IMBIE assessment¹ and to evaluate an updated definition tailored towards mass budget assessments. The first set comprises 19 drainage basins delineated using surface elevation maps derived from ICESat-1 with a total area of 1,703,625 km² (ref. ²⁰). The second drainage basin set is an updated definition that considers other factors such as the direction of ice flow and includes 6 basins with a combined area of 1,723,300 km² (ref. ³⁷). The two drainage basin sets differ by 1% in area at the scale of the Greenland Ice Sheet, and this has a negligible impact on mass trends when compared to the estimated uncertainty of individual techniques.

Glacial isostatic adjustment

GIA (the delayed response of Earth's interior to temporal changes in ice loading) affects estimates of ice sheet mass balance determined from satellite gravimetry and, to a lesser extent, satellite altimetry⁹³. Here, we compare six independent models of GIA in the vicinity of the Greenland Ice Sheet (Extended Data Table 1). The GIA model solutions we considered differ for a variety of reasons, including differences in their physics, in their computational approach, in their prescriptions of solid Earth unloading during the last glacial cycle and their Earth rheology, and in the datasets against which they are evaluated. Although alternative ice histories (for example, ref. ⁹⁴) and mantle viscosities (for example, ref. ⁹⁵) are available, we restricted our comparison to those that contributed to our assessment. No approach is generally accepted as optimal, and so we evaluate the models by computing the mean and standard deviation of their predicted uplift rates (Extended Data Fig. 3). We also estimate the contribution of each model to gravimetric mass trends using a common processing approach⁴¹ that puts special emphasis on the treatment of low spherical harmonic degrees in the GIA-related trends in the gravitational field.

The highest rates of GIA-related uplift occur in northern Greenland, although this region also exhibits marked variability among the solutions, as does the area around Kangerlussuaq Glacier to the southeast. Even though the model spread is high in northern Greenland, the signal in this sector is also consistently high in most solutions. However, none of the GIA models considered here fully captures all areas of high uplift present in the models, and so it is possible there is a bias towards low values in the average field across the ice sheet overall. The models yield an average adjustment for GRACE estimates of the Greenland Ice Sheet mass balance of -3 Gt yr^{-1} , with a standard deviation of around 20 Gt yr^{-1} . The spread is probably due, in part, to differences in the way each model accounts for GIA in North America (which is ongoing and impacts western Greenland), and so care must be taken when estimating mass balance at the basin scale. Local misrepresentation of the solid Earth response can also have a relatively large impact stemming

especially from lateral variations of solid Earth properties^{42,51}, and revisions of the current state of knowledge can be expected³⁴.

SMB

Here, ice-sheet SMB is defined as total precipitation minus sublimation, evaporation and meltwater runoff; that is, the interaction of the atmosphere and the superficial snow and firn layers, for example through mass exchanges via precipitation, sublimation and runoff, and through mass redistribution by snowdrift, melting and refreezing. We compare ten estimates of Greenland Ice Sheet SMB derived using a range of alternative approaches; four regional climate models (RCMs), two downscaled RCMs, a global reanalysis, two downscaled model reanalyses of climate data and one gridded model of snow processes driven by climate model output (Extended Data Table 2).

Although SMB models of similar classes tend to produce similar results, there are larger differences between classes—most notably the global reanalysis and the process model—which lead to estimates of SMB that are substantially higher and lower than all other solutions, respectively. The regional climate model solutions agree well at the scale of individual drainage sectors, with the largest differences occurring in northeast Greenland (Extended Data Fig. 4). The snow process model tends to underestimate SMB when compared with the other solutions that we have considered in various sectors of the ice sheet, at times even yielding negative SMB, while the global reanalysis tends to overestimate it.

Across all models, the average SMB of the Greenland Ice Sheet between 1980 and 2012 is 351 Gt yr^{-1} and the standard deviation is 98 Gt yr^{-1} . However, the spread among the 8 RCM and downscaled reanalyses is considerably smaller; these solutions lead to an average Greenland Ice Sheet SMB of 361 Gt yr^{-1} with a standard deviation of 40 Gt yr^{-1} over the same period. By comparison, the global reanalysis and process model lead to ice-sheet-wide estimates of SMB that are considerably larger (504 Gt yr^{-1}) and smaller (125 Gt yr^{-1}) than this range, respectively. Model resolution is an important factor when estimating SMB and its components, as respective contributions where only the spatial resolution differed yield regional differences. The underlying model domains were also identified as a source of discrepancy in the case of the Greenland Ice Sheet, as some products would allocate the ablation area outside the given mask.

Individual estimates of ice sheet mass balance

To standardize our comparison and aggregation of the 26 individual satellite estimates of Greenland Ice Sheet mass balance, we applied a common approach to derive rates of mass change from cumulative mass trends⁴¹. Rates of mass change were computed over 36-month intervals centred on regularly spaced (monthly) epochs within each cumulative mass trend time series, oversampling the individual time series where necessary. At each epoch, rates of mass change were estimated by fitting a linear trend to data within the surrounding 36-month time window using a weighted least-squares approach, with each point weighted by its measurement error. The associated mass trend uncertainties were estimated as the root sum square of the regression error and the measurement error. Time series were truncated by half the moving-average window period at the start and end of their period. The emerging rates of mass change were then averaged over calendar years to reduce the impact of seasonal cycles.

Gravimetry. We include 14 estimates of Greenland Ice Sheet ice sheet mass balance determined from GRACE satellite gravimetry that together span the period 2003–2016 (Extended Data Fig. 1). Ten of the gravimetry solutions were computed using spherical harmonic solutions to the global gravity field and four were computed using spatially defined mass concentration units (Supplementary Table 1). An unrestricted range of alternative GIA corrections were used in the formation of the gravimetry mass balance solutions based on commonly adopted

model solutions and their variants^{34,51–57} (Supplementary Table 1). All of the gravimetry mass balance solutions included in this study use the same degree-1 coefficients to account for geocentre motion⁵⁸ and, although an alternative set is now available⁹⁶, the estimated improvement in certainty is small in comparison with their magnitude and spread. There was some variation in the sampling of the individual gravimetry datasets, and their collective effective (weighted mean) temporal resolution is 0.08 yr. Overall, there is good agreement between rates of Greenland Ice Sheet mass change derived from satellite gravimetry (Extended Data Fig. 5); all solutions show the ice sheet to be in a state of negative mass balance throughout their survey periods, with mass loss peaking in 2011 and reducing thereafter. During the period 2005–2015, annual rates of mass change determined from satellite gravimetry differ by 104 Gt yr⁻¹ on average, and their average standard deviation is 31 Gt yr⁻¹ (Extended Data Table 3).

Altimetry. We include nine estimates of Greenland Ice Sheet mass balance determined from satellite altimetry that together span the period 2004–2018 (Extended Data Fig. 1). Three of the solutions are derived from radar altimetry, four from laser altimetry and two use a combination of both (Supplementary Table 1). The altimetry mass trends are also computed using a range of approaches, including crossovers, planar fits and repeat track analyses. The laser altimetry mass trends are computed from ICESat-1 data as constant rates of mass change over their respective survey periods, whereas the radar altimetry mass trends are computed from EnviSat and/or CryoSat-2 data with a temporal resolution of between 1 and 72 months. In consequence, the altimetry solutions have an effective collective temporal resolution of 0.74 yr. Mass changes are computed after making corrections for alternative sources of surface elevation change, including glacial isostatic and elastic adjustment, and firn height changes (see Supplementary Table 1). Despite the range of input data and technical approaches, there is good overall agreement between rates of mass change determined from the various satellite altimetry solutions (Extended Data Fig. 5). All altimetry solutions show the Greenland Ice Sheet to be in a state of negative mass balance throughout their survey periods, with mass loss peaking in 2012 and reducing thereafter. During the period 2005–2015, annual rates of mass change determined from satellite altimetry differ by 121 Gt yr⁻¹ on average, and their average standard deviation is 42 Gt yr⁻¹ (Extended Data Table 3). The greatest variance lies among the 4 laser altimetry mass balance solutions, which range from –248 to –128 Gt yr⁻¹ between 2004 and 2010; aside from methodological differences; possible explanations for this high spread include the relatively short period over which the mass trends are determined, the poor temporal resolution of these datasets and the rapid change in mass balance occurring during the period in question.

IOM. We include three estimates of Greenland Ice Sheet mass balance determined from the IOM that together span the period 1992–2015 (Extended Data Fig. 1). Although there are relatively few datasets in comparison with the gravimetry and altimetry solutions, the input–output data provide information on the partitioning of the mass change (surface processes and/or ice dynamics) cover a considerably longer period and are therefore an important record of changes in Greenland Ice Sheet mass during the 1990s. The IOM makes use of a wide range of satellite imagery (for example, refs. ^{6,40,97–102}) combined with measurements of ice thickness (for example, ref. ¹⁰³) for computing ice sheet discharge (output), and several alternative SMB model estimates of snow accumulation (input) and runoff (output) (see Supplementary Table 1). Two of the IOM datasets exhibit temporal variability across their survey periods, and two provide only constant rates of mass changes. Although these latter records are relatively short, they are an important marker with which variances among independent estimates can be evaluated. The collective effective (weighted mean) temporal resolution of the IOM data are 0.14 yr, although it should be

noted that in earlier years the satellite ice discharge component of the data are relatively sparsely sampled in time (for example, ref. ¹⁰⁴). There is good overall agreement between rates of mass change determined from the input–output method solutions (Extended Data Fig. 5). During the period 2005–2015, annual rates of mass change determined from the four input–output datasets differ by up to 48 Gt yr⁻¹ on average, and their average standard deviation is 23 Gt yr⁻¹ (Extended Data Table 3). These differences are comparable to the estimated uncertainty of the individual techniques and are also small relative to the estimated mass balance over the period in question. In addition to showing that the Greenland Ice Sheet was in a state of negative mass balance since 2000, with mass loss peaking in 2012 and reducing thereafter, the IOM data show that the ice sheet was close to a state of balance before this period³³.

Aggregate estimate of ice sheet mass balance

To produce an aggregate estimate of Greenland Ice Sheet mass balance, we combine the 14 gravimetry, 9 altimetry and 3 IOM datasets to produce a single 26-yr record spanning the period 1992–2018. First, we combine the gravimetry, altimetry and the IOM data separately into three monthly time series by forming an error-weighted average of individual monthly rates of ice sheet mass change computed using the same technique (Extended Data Fig. 6). At each epoch, we estimate the uncertainty of these time-series as the root mean square of their component time-series errors. We then combine the mass balance time series derived from gravimetry, altimetry and the IOM to produce a single aggregate (reconciled) estimate, computed as the error-weighted mean of mass trends sampled at each epoch. We estimated the uncertainty of this reconciled rate of mass balance as either the root mean square departure of the constituent mass trends from their weighted-mean or the root mean square of their uncertainties, whichever is larger. Cumulative uncertainties are computed as the root sum square of annual errors, on the assumption that annual errors are not correlated over time. This assumption has been employed in numerous mass balance studies^{1,17,33,41}, and its effect is to reduce cumulative errors by a factor of 2.2 over the 5-yr periods we employ in this study (Table 1). If some sources of error are temporally correlated, the cumulative uncertainty may therefore be underestimated. In a recent study, for example, it is estimated that 30% of the annual mass balance error is systematic¹⁰⁵, and in this instance the cumulative error may be 37% larger. On the other hand, the estimated annual error on aggregate mass trends reported in this study (61 Gt yr⁻¹) are 70% larger than the spread of the independent estimates from which they are combined (36 Gt yr⁻¹) (Extended Data Table 3), which suggests the underlying errors may be overestimated by a similar degree. A more detailed analysis of the measurement and systematic errors is required to improve the cumulative error budget.

During the period 2004–2015, when all three satellite techniques were in operation, there is good agreement between changes in ice sheet mass balance on a variety of timescales (Extended Data Fig. 6). In Greenland, there are large annual cycles in mass superimposed on equally prominent interannual fluctuations as well as variations of intermediate (~5 yr) duration. These signals are consistent with fluctuations in SMB that have been identified in meteorological records^{1,59}, and are present within the time series of mass balance emerging from all three satellite techniques, to varying degrees, according to their effective temporal resolution. For example, correlated seasonal cycles are apparent in the gravimetry and IOM mass balance time series, because their effective temporal resolutions are sufficiently short (0.08 and 0.14 yr, respectively) to resolve such changes. However, at 0.74 yr, the effective temporal resolution of the altimetry mass balance time series is too coarse to detect cycles on sub-annual timescales. Nevertheless, when the aggregated mass balance data emerging from all three experiment groups are degraded to a common temporal resolution of 36 months, the time series are well correlated ($0.63 < r^2 < 0.80$) and, over longer periods, all techniques identify the marked increases in Greenland

Article

Ice Sheet mass loss peaking in 2012. During the period 2005–2015, annual rates of mass change determined from all three techniques differ by up to 162 Gt yr^{-1} on average, and their average standard deviation is 41 Gt yr^{-1} —a value that is small when compared with their estimated uncertainty (18 Gt yr^{-1}) (Extended Data Table 3).

Data availability

The aggregated Greenland Ice Sheet mass balance data and estimated errors generated in this study are freely available at <http://imbie.org> and at the NERC Polar Data Centre, <https://doi.org/10.5285/8D5FF221-A470-4CC1-B7C4-CBDF383554FC>.

Code availability

The code used to compute and aggregate rates of ice sheet mass change and their estimated errors are freely available at <https://github.com/IMBIE>.

51. Peltier, W. R., Argus, D. F. & Drummond, R. Space geodesy constrains ice age terminal deglaciation: the global ICE-6G_C (VM5a) model. *J. Geophys. Res. Solid Earth* **120**, 450–487 (2015).
52. Paulson, A., Zhong, S. & Wahr, J. Inference of mantle viscosity from GRACE and relative sea level data. *Geophys. J. Int.* **171**, 497–508 (2007).
53. Peltier, W. R. Global glacial isostasy and the surface of the Ice-Age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
54. Simpson, M. J. R., Milne, G. A., Huybrechts, P. & Long, A. J. Calibrating a glaciological model of the Greenland ice sheet from the Last Glacial Maximum to present-day using field observations of relative sea level and ice extent. *Quat. Sci. Rev.* **28**, 1631–1657 (2009).
55. A. G., Wahr, J. & Zhong, S. Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to glacial isostatic adjustment in Antarctica and Canada. *Geophys. J. Int.* **192**, 557–572 (2013).
56. Schrama, E. J. O., Wouters, B. & Rietbroek, R. A mascon approach to assess ice sheet and glacier mass balances and their uncertainties from GRACE data. *J. Geophys. Res. Solid Earth* **119**, 6048–6066 (2014).
57. Klemann, V. & Martinec, Z. Contribution of glacial-isostatic adjustment to the geocenter motion. *Tectonophysics* **511**, 99–108 (2011).
58. Swenson, S., Chambers, D. & Wahr, J. Estimating geocenter variations from a combination of GRACE and ocean model output. *J. Geophys. Res. Solid Earth* **113**, B08410 (2008).
59. Wouters, B., Bamber, J. L., van den Broeke, M. R., Lenaerts, J. T. M. & Sasgen, I. Limits in detecting acceleration of ice sheet mass loss due to climate variability. *Nat. Geosci.* **6**, 613–616 (2013).
60. Bonin, J. & Chambers, D. Uncertainty estimates of a GRACE inversion modelling technique over Greenland using a simulation. *Geophys. J. Int.* **194**, 212–229 (2013).
61. Blazquez, A. et al. Exploring the uncertainty in GRACE estimates of the mass redistributions at the Earth surface: implications for the global water and sea level budgets. *Geophys. J. Int.* **215**, 415–430 (2018).
62. Forsberg, R., Sørensen, L. & Simonsen, S. Greenland and Antarctica Ice Sheet Mass Changes and Effects on Global Sea Level. *Surv. Geophys.* **38**, 89–104 (2017).
63. Groh, A. & Horwath, M. The method of tailored sensitivity kernels for GRACE mass change estimates. *Geophys. Res. Abstr.* **18**, 12065 (2016).
64. Harig, C. & Simons, F. J. Mapping Greenland's mass loss in space and time. *Proc. Natl Acad. Sci. USA* **109**, 19934–19937 (2012).
65. Luthcke, S. B. et al. Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *J. Glaciol.* **59**, 613–631 (2013).
66. Andrews, S. B., Moore, P. & King, M. A. Mass change from GRACE: a simulated comparison of Level-1B analysis techniques. *Geophys. J. Int.* **200**, 503–518 (2015).
67. Save, H., Bettadpur, S. & Tapley, B. D. High-resolution CSR GRACE RLO5 mascons. *J. Geophys. Res. Solid Earth* **121**, 7547–7569 (2016).
68. Seo, K.-W. et al. Surface mass balance contributions to acceleration of Antarctic ice mass loss during 2003–2013. *J. Geophys. Res. Solid Earth* **120**, 3617–3627 (2015).
69. Velicogna, I., Sutterley, T. C. & van den Broeke, M. R. Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data. *Geophys. Res. Lett.* **41**, 8130–8137 (2014).
70. Vishwakarma, B. D., Horwath, M., Devaraju, B., Groh, A. & Sneeuw, N. A data-driven approach for repairing the hydrological catchment signal damage due to filtering of GRACE products. *Wat. Resour. Res.* **53**, 9824–9844 (2017).
71. Wiese, D. N., Landerer, F. W. & Watkins, M. M. Quantifying and reducing leakage errors in the JPL RLO5M GRACE mascon solution. *Wat. Resour. Res.* **52**, 7490–7502 (2016).
72. Ivins, E. R. & James, T. S. Antarctic glacial isostatic adjustment: a new assessment. *Antarct. Sci.* **17**, 541–553 (2005).
73. Ivins, E. R. et al. Antarctic contribution to sea level rise observed by GRACE with improved GIA correction. *J. Geophys. Res. Solid Earth* **118**, 3126–3141 (2013).
74. Rodell, M. et al. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
75. Döll, P., Kaspar, F. & Lehner, B. A global hydrological model for deriving water availability indicators: model tuning and validation. *J. Hydrol.* **270**, 105–134 (2003).
76. Cheng, M., Tapley, B. D. & Ries, J. C. Deceleration in the Earth's oblateness. *J. Geophys. Res. Solid Earth* **118**, 740–747 (2013).
77. Balmaseda, M. A., Mogensen, K. & Weaver, A. T. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.* **139**, 1132–1161 (2013).
78. Pujol, M.-I. et al. DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years. *Ocean Sci.* **12**, 1067–1090 (2016).
79. Menemenlis, D. et al. ECCO2: High resolution global ocean and sea ice data synthesis. In *AGU Fall Meeting Abstracts 2008* OS31C-1292 (AGU, 2008).
80. Dobslaw, H. et al. Simulating high-frequency atmosphere-ocean mass variability for dealiasing of satellite gravity observations: AOD1B RLO5. *J. Geophys. Res. Oceans* **118**, 3704–3711 (2013).
81. Carrère, L. & Lyard, F. Modeling the barotropic response of the global ocean to atmospheric wind and pressure forcing – comparisons with observations. *Geophys. Res. Lett.* **30**, 1275 (2003).
82. Csatho, B. M. et al. Laser altimetry reveals complex pattern of Greenland Ice Sheet dynamics. *Proc. Natl Acad. Sci. USA* **111**, 18478–18483 (2014).
83. Nilsson, J., Gardner, A., Sandberg Sørensen, L. & Forsberg, R. Improved retrieval of land ice topography from CryoSat-2 data and its impact for volume-change estimation of the Greenland Ice Sheet. *Cryosphere* **10**, 2953–2969 (2016).
84. Gourmelen, N. et al. CryoSat-2 swath interferometric altimetry for mapping ice elevation and elevation change. *Adv. Space Res.* **62**, 1226–1242 (2018).
85. Gunter, B. C. et al. Empirical estimation of present-day Antarctic glacial isostatic adjustment and ice mass change. *Cryosphere* **8**, 743–760 (2014).
86. Helm, V., Humbert, A. & Miller, H. Elevation and elevation change of Greenland and Antarctica derived from CryoSat-2. *Cryosphere* **8**, 1539–1559 (2014).
87. Kjeldsen, K. K. et al. Improved ice loss estimate of the northwestern Greenland ice sheet. *J. Geophys. Res. Solid Earth* **118**, 698–708 (2013).
88. Felikson, D. et al. Comparison of elevation change detection methods from ICESat altimetry over the Greenland Ice Sheet. *IEEE Trans. Geosci. Remote Sens.* **55**, 5494–5505 (2017).
89. Andersen, M. L. et al. Basin-scale partitioning of Greenland ice sheet mass balance components (2007–2011). *Earth Planet. Sci. Lett.* **409**, 89–95 (2015).
90. Colgan, W. et al. Greenland ice sheet mass balance assessed by PROMICE (1995–2015). *Geol. Surv. Denmark Greenl. Bull.* **43**, e2019430201 (2019).
91. van Wessem, J. M. et al. Updated cloud physics in a regional atmospheric climate model improves the modelled surface energy balance of Antarctica. *Cryosphere* **8**, 125–135 (2014).
92. Fettweis, X. et al. Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR. *Cryosphere* **7**, 469–489 (2013).
93. Wahr, J., Wingham, D. & Bentley, C. A method of combining ICESat and GRACE satellite data to constrain Antarctic mass balance. *J. Geophys. Res. Solid Earth* **105**, 16279–16294 (2000).
94. Lambeck, K., Roubey, H., Purcell, A., Sun, Y. & Sambridge, M. Closing the sea level budget at the Last Glacial Maximum. *Proc. Natl Acad. Sci. USA* **111**, 15861–15862 (2014).
95. Caron, L., Métivier, L., Greff-Lefftz, M., Fleitout, L. & Roubey, H. Inverting Glacial Isostatic Adjustment signal using Bayesian framework and two linearly relaxing rheologies. *Geophys. J. Int.* **209**, 1126–1147 (2017).
96. Sun, Y., Riva, R. & Ditmar, P. Optimizing estimates of annual variations and trends in geocenter motion and J2 from a combination of GRACE data and geophysical models. *J. Geophys. Res. Solid Earth* **121**, 8352–8370 (2016).
97. Nagler, T., Rott, H., Hetzenecker, M., Wuite, J. & Potin, P. The Sentinel-1 Mission: New Opportunities for Ice Sheet Observations. *Remote Sens.* **7**, 9371–9389 (2015).
98. Mouginot, J., Rignot, E., Scheuchl, B. & Millan, R. Comprehensive annual ice sheet velocity mapping using Landsat-8, Sentinel-1, and RADARSAT-2 data. *Remote Sens.* **9**, 364 (2017).
99. Joughin, I., Smith, B. E. & Howat, I. Greenland Ice Mapping Project: ice flow velocity variation at sub-monthly to decadal timescales. *Cryosphere* **12**, 2211–2227 (2018).
100. Lemos, A. et al. Ice velocity of Jakobshavn Isbræ, Petermann Glacier, Nioghalvfjærdsfjorden, and Zachariæ Isstrøm, 2015–2017, from Sentinel 1-a/b SAR imagery. *Cryosphere* **12**, 2087–2097 (2018).
101. Joughin, I. et al. Continued evolution of Jakobshavn Isbræ following its rapid speedup. *J. Geophys. Res. Earth Surf.* **113**, F04006 (2008).
102. Joughin, I., Abdalati, W. & Fahnestock, M. Large fluctuations in speed on Greenland's Jakobshavn Isbræ glacier. *Nature* **432**, 608–610 (2004).
103. Gogineni, S. et al. Coherent radar ice thickness measurements over the Greenland ice sheet. *J. Geophys. Res. D Atmospheres* **106**, 33761–33772 (2001).
104. Rignot, E. et al. Recent Antarctic ice mass loss from radar interferometry and regional climate modelling. *Nat. Geosci.* **1**, 106–110 (2008).
105. Shepherd, A. et al. Trends in Antarctic Ice Sheet elevation and mass. *Geophys. Res. Lett.* **46**, 8174–8183 (2019).
106. Martinec, Z. & Hagedoorn, J. The rotational feedback on linear-momentum balance in glacial isostatic adjustment. *Geophys. J. Int.* **199**, 1823–1846 (2014).
107. Fretwell, P. et al. Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7**, 375–393 (2013).
108. Rignot, E., Mouginot, J. & Scheuchl, B. Ice flow of the Antarctic Ice Sheet. *Science* **333**, 1427–1430 (2011).
109. Rignot, E., Mouginot, J. & Scheuchl, B. Antarctic grounding line mapping from differential satellite radar interferometry. *Geophys. Res. Lett.* **38**, L10504 (2011).
110. Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H. & Box, J. E. Liquid water flow and retention on the Greenland Ice Sheet in the regional climate model HIRHAM5: local and large-scale impacts. *Front. Earth Sci.* **4**, 110 (2017).
111. Martinec, Z. Spectral-finite element approach to three-dimensional viscoelastic relaxation in a spherical earth. *Geophys. J. Int.* **142**, 117–141 (2000).
112. Fleming, K. & Lambeck, K. Constraints on the Greenland Ice Sheet since the Last Glacial Maximum from sea-level observations and glacial-rebound models. *Quat. Sci. Rev.* **23**, 1053–1077 (2004).

113. King, M. A., Whitehouse, P. L. & van der Wal, W. Incomplete separability of Antarctic plate rotation from glacial isostatic adjustment deformation within geodetic observations. *Geophys. J. Int.* **204**, 324–330 (2016).
114. Spada, G., Melini, D. & Colleoni, F. SELEN v2.9.12 (Computational Infrastructure for Geodynamics, 2018); <https://geodynamics.org/cig/software/selen>.
115. Noël, B. et al. Evaluation of the updated regional climate model RACMO2.3: summer snowfall impact on the Greenland Ice Sheet. *Cryosphere* **9**, 1831–1844 (2015).
116. Noël, B. et al. A daily, 1 km resolution data set of downscaled Greenland ice sheet surface mass balance (1958–2015). *Cryosphere* **10**, 2361–2377 (2016).
117. Gelaro, R. et al. The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
118. Wilton, D. J. et al. High resolution (1 km) positive degree-day modelling of Greenland ice sheet surface mass balance, 1870–2012 using reanalysis data. *J. Glaciol.* **63**, 176–193 (2017).
119. Mernild, S. H., Liston, G. E., Hiemstra, C. A. & Christensen, J. H. Greenland Ice Sheet surface mass-balance modeling in a 131-yr perspective, 1950–2080. *J. Hydrometeorol.* **11**, 3–25 (2010).

Acknowledgements This work is an outcome of the IMBIE supported by the ESA Climate Change Initiative and the NASA Cryosphere Program. A.S. was additionally supported by a Royal Society Wolfson Research Merit Award and the UK Natural Environment Research Council Centre for Polar Observation and Modelling.

Author contributions A.S. and E.I. designed and led the study. E.R., B.S., M.v.d.B., I.V. and P.W. led the IOM, altimetry, SMB, gravimetry and GIA experiments, respectively. G.K., S.N.,

T.P. and T. Scambos provided additional supervision on glaciology. K.B., A.H., I.J., M.E.E. and T.W. provided additional supervision on satellite observations and N.S. provided additional supervision on GIA. G.M., M.E.P. and T. Slater performed the mass balance data collation and analysis. T. Slater performed the AR5 data analysis. P.W. and I.S. performed the GIA data analysis. M.v.W. and T. Slater performed the SMB data analysis. A.S., E.I., K.B., M.E., N.G., A.H., H.K., M.M., I.O., I.S., T. Slater, M.v.W. and P.W. wrote the manuscript. A.S. led the writing, E.I., K.B., M.E., and T. Slater led the drafting and editing, M.v.W. led the SMB text, P.W. and I.S. led the GIA text and N.G., A.H., H.K., M.M. and I.O. contributed elsewhere. A.S., K.B., H.K., G.M., M.E.P., I.S., S.B.S., T. Slater, P.W. and M.v.W. prepared the figures and tables, with particular focus on Fig. 1 (S.B.S), Fig. 3 (T. Slater), Fig. 4 (T. Slater), Extended Data Fig. 2 (K.B.), Extended Data Fig. 3 (P.W.), Extended Data Fig. 2 (M.v.W.), Extended Data Table 1 (P.W. and I.S.), Extended Data Table 2 (M.v.W.) and Supplementary Table 1 (H.K. and T. Slater). G.M. and M.E.P. led the production of all other figures and tables. All authors participated in the data interpretation and commented on the manuscript.

Competing interests The authors declare no competing interests.

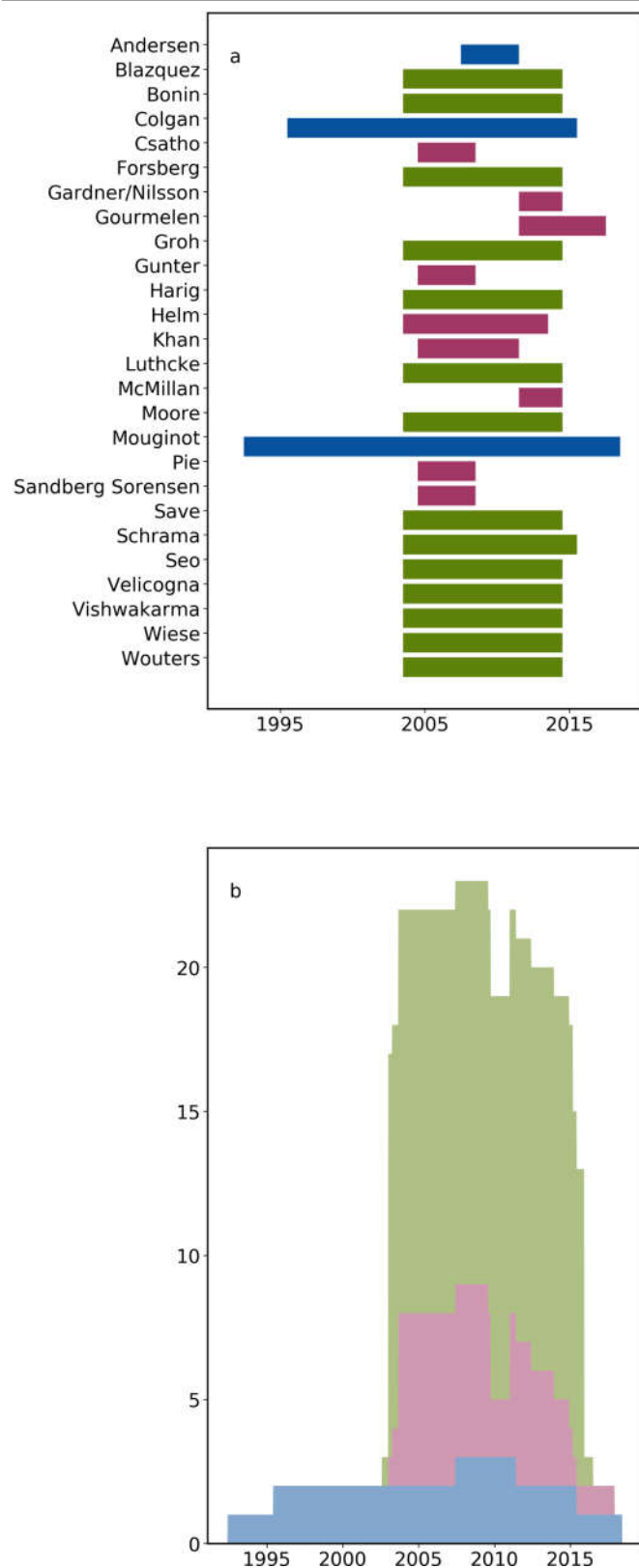
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1855-2>.

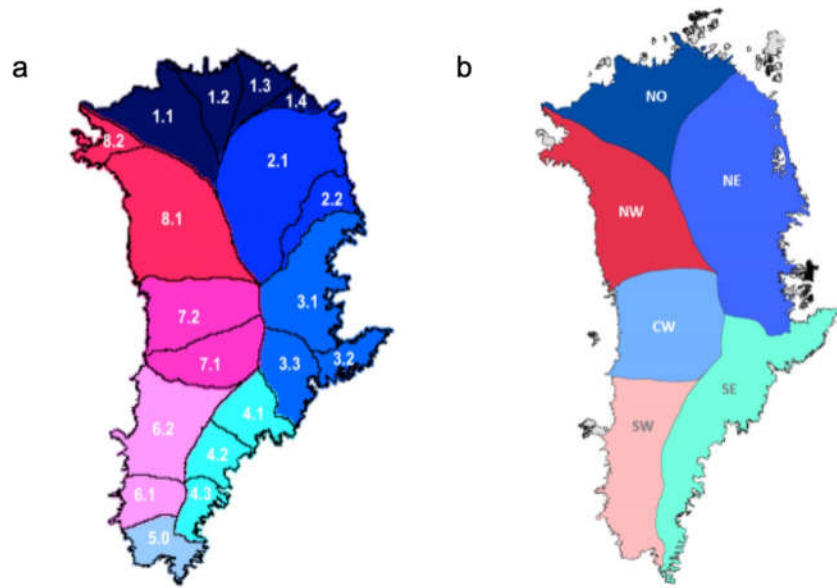
Correspondence and requests for materials should be addressed to A.S.

Peer review information *Nature* thanks Christina Hulbe, Andreas Kääb and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

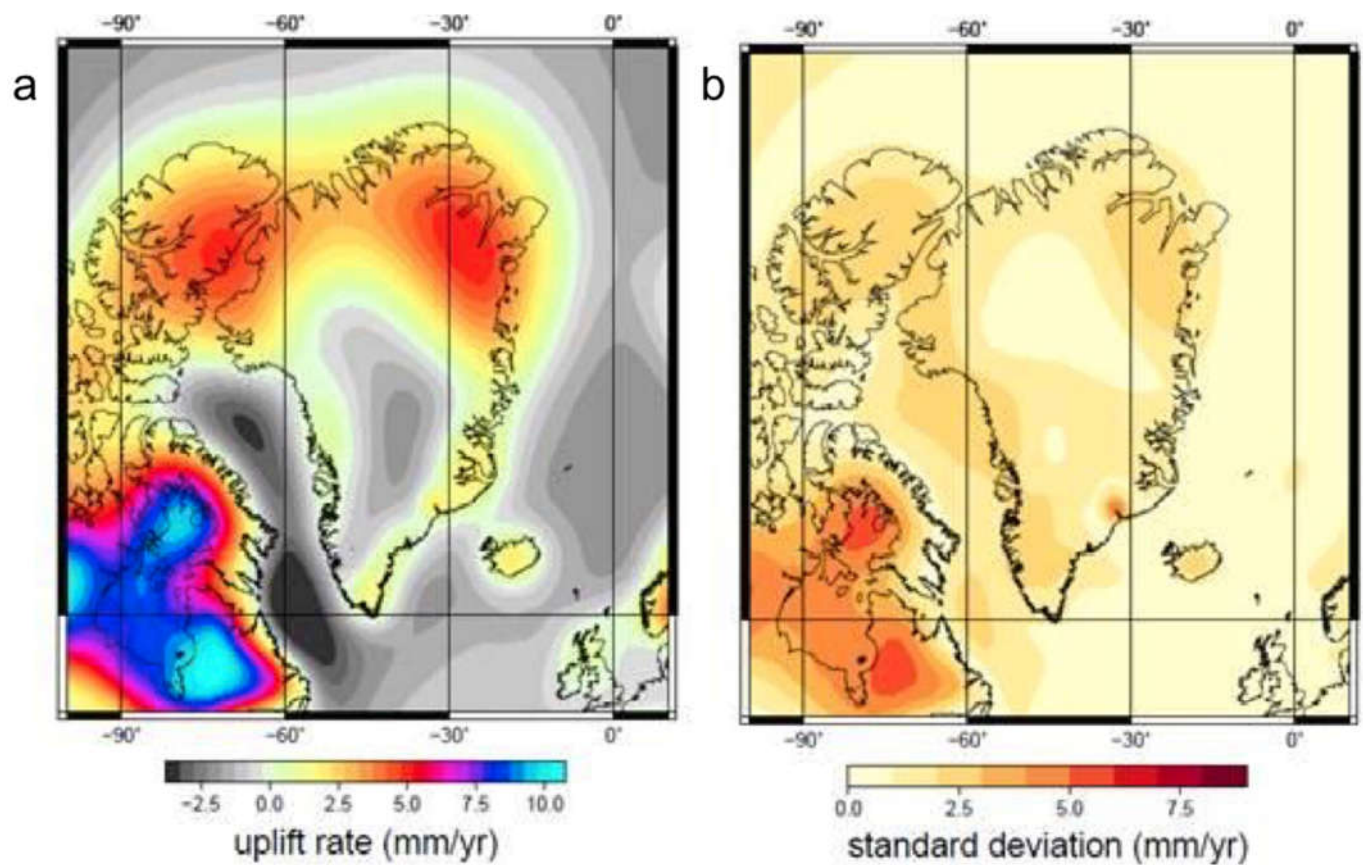
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Ice sheet mass balance datasets. a, Participant datasets used in this study and their main contributors. **b,** The number of data available in each calendar year. The interval 2003–2010 includes almost all datasets and is selected as the overlap period. Further details of the satellite observations used in this study are provided in Supplementary Table 1. Refs. ^{28, 33, 38, 56, 59–71, 82–90}.

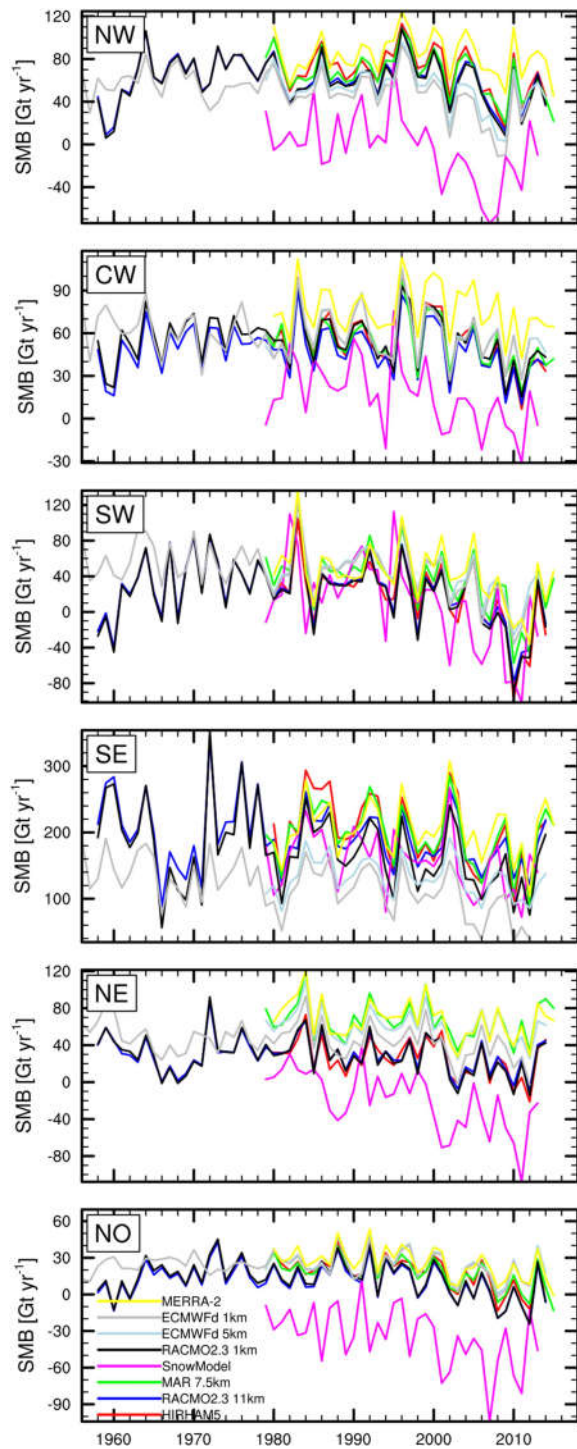


Extended Data Fig. 2 | Greenland Ice Sheet drainage basins. a, b, Basin used in this study, according to the definitions of ref. ²⁰ (a) and ref. ³⁷ (b).

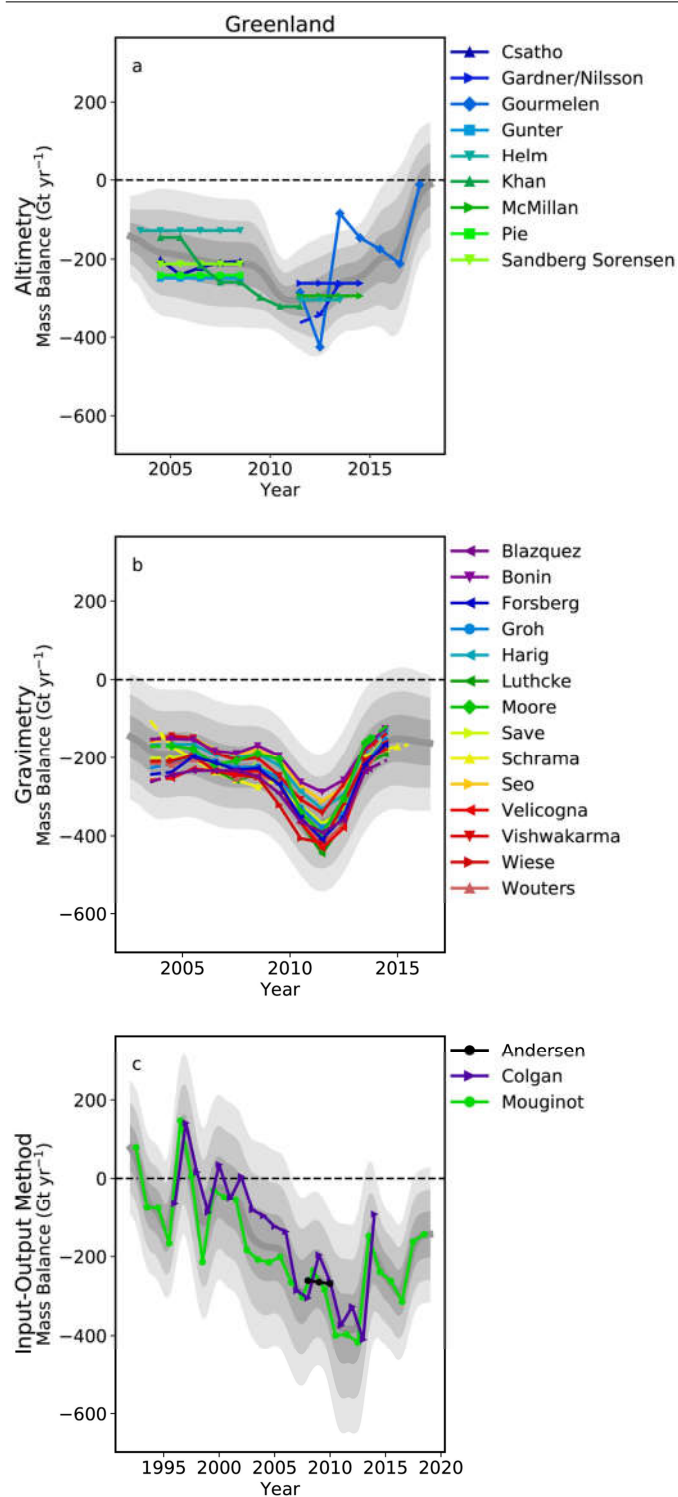


Extended Data Fig. 3 | Modelled glacial isostatic adjustment in Greenland. a, Bedrock uplift rates in Greenland averaged over the GIA model solutions used in this study **(a)** and their standard deviation **(b)**. Further details of the GIA

models used in this study are provided in Extended Data Table 1. High rates of uplift and subsidence associated with the former Laurentide Ice Sheet are apparent to the southwest of Greenland.

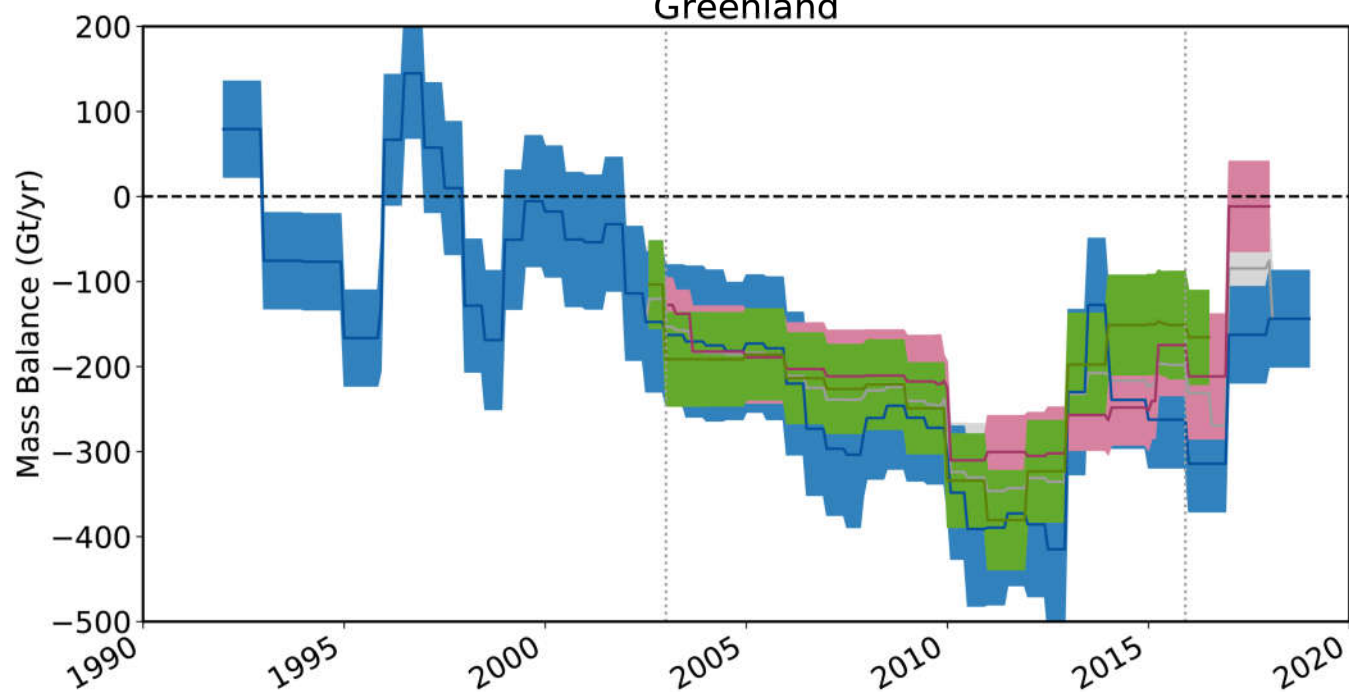


Extended Data Fig. 4 | SMB of the Greenland Ice Sheet. a–f, Time series of SMB in the NW, CW, SW, SE, NE and NO Greenland Ice Sheet drainage basins (Extended Data Fig. 2)^{108,109}. Solid lines are annual averages of the monthly data (dashed lines). Further details of the SMB models used in this study are provided in Extended Data Table 2.



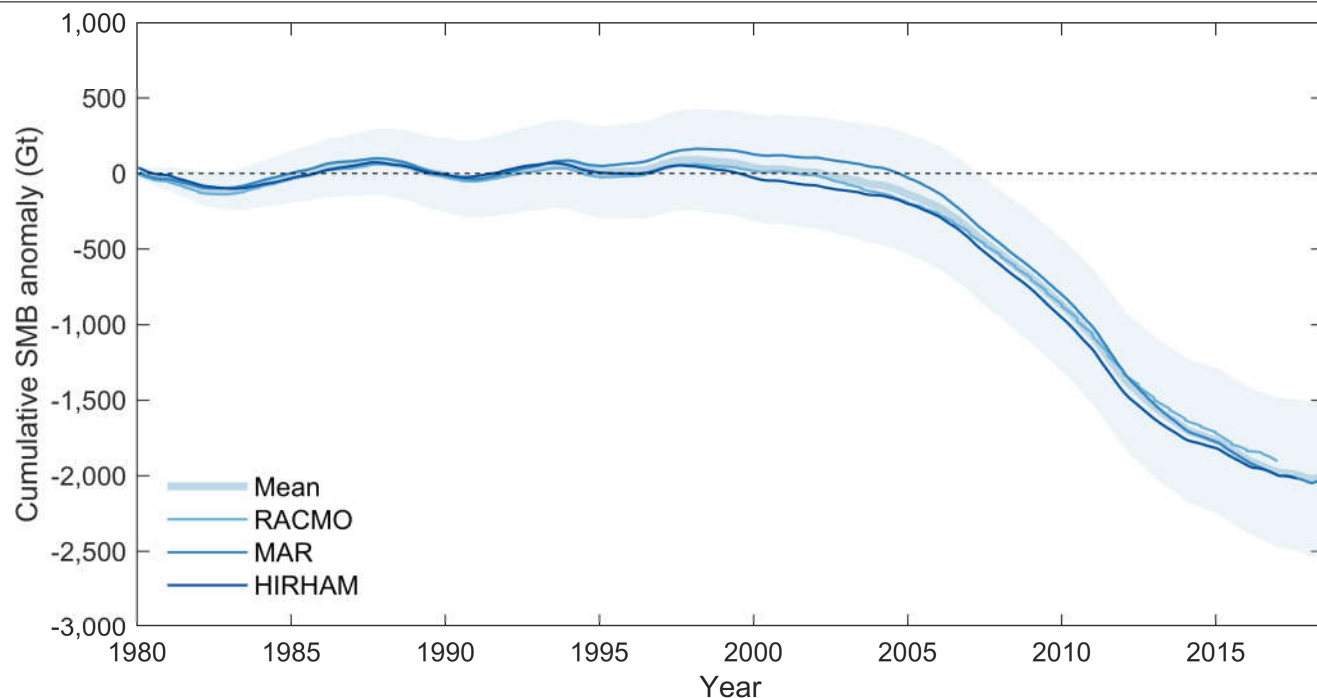
Extended Data Fig. 5 | Greenland Ice Sheet mass balance intracomparison. **a–c**, Individual rates of Greenland Ice Sheet mass balance used in this study as determined from satellite altimetry (**a**), gravimetry (**b**) and the input–output method (**c**). The grey shading shows the estimated 1σ (dark), 2σ (mid-) and 3σ (light) uncertainty relative to the ensemble average. Refs. ^{28,33,38,56,59–71,82–90}.

dM/dt intercomparison Greenland



Extended Data Fig. 6 | Greenland Ice Sheet mass balance intercomparison.
Rate of Greenland Ice Sheet mass balance as derived from the three techniques: satellite radar and laser altimetry (red), input-output method (blue) and

gravimetry (green). Their arithmetic mean is shown in grey. The estimated uncertainty is also shown (shaded envelopes) and is computed as the root mean square of the component time-series errors.



Extended Data Fig. 7 | Cumulative Greenland Ice Sheet SMB. The cumulative surface mass change determined from an average (mean) of the RACMO2.3p2⁴⁶, MARv3.6²¹ and HIRHAM⁹ regional climate models relative to their 1980–1990 means (see Methods). The estimated uncertainty of the mean change is also shown (shaded area), computed as the average of the uncertainties from each of the three models. RACMO2.3p2 uncertainties are

based on a comparison to in situ observations³³. MARv3.6 uncertainties are evaluated from the variability due to forcing from climate reanalyses²¹. HIRHAM uncertainties are estimated on the basis of comparisons to in situ accumulation and ablation data¹⁰. Cumulative uncertainties are computed as the root sum square of annual errors, on the assumption that these errors are not correlated over time¹⁷.

Extended Data Table 1 | Details of GIA models used in this study

Contributor	Model	Publication ^a	Earth model ^b	Ice model ^b	GIA model ^c	Constraint data ^d	GIA (Gt/yr)
A	A13	A et al., 2013	VM5a (1D) ^e	ICE-6G_C ^f	SH, C, RF, SG, OL	As for ICE-6G_C ^f	-9 [‡]
Lecavalier	Huy3	Lecavalier et al., 2014	1D (120, 0.5, 2)	Huy3/ICE-5G	SH(256), IC, RF, SG, OL	RSL, ice extent, paleo thinning rates	-19 [‡]
Sasgen	GGG1D.0	Khan et al., 2016 Martinec, 2000	VM-GPS ⁴²	modified GREEN1 ⁷⁹	SH(256)/F E(radial), IC, RF, SG, OL	GPS, RSL	+17 [†]
Peltier	ICE-6G_D (VM5a)	Peltier et al., 2015	VM5a (1D) ^e	ICE-6G_D ^g	SH(512)	GPS, RSL, Earth rotation	-10 [‡]
van der Wal	SL-dry-4mm/W 12	King et al., 2016	3D, power-law rheology	Combination of W12 (Antarctica) and ICE-5G	FE, IC, xRF	GPS, RSL, seismic velocities (Earth model)	+21 [†]
Spada	SELEN 4	Spada et al., 2018	VM5a (3-layer average of 1D model) ^e	ICE-6G_C ^f	SELEN4: SH(128), IC, RF, SG, OL	As for ICE-6G_C ^f	-27 [‡]

Refs. ^{34,42,51,55,111,113,114}.

[†]Regional changes in mass associated with the GIA signal determined by the contributor.

[‡]Regional changes in mass associated with the GIA signal calculated as an indicative rate using spherical-harmonic degrees 3 to 90 and a common treatment of degree 2 (ref. ¹⁰⁶).

^aMain reference publication(s).

^bModel from main publication unless otherwise stated. Comma-separated values refer to properties of a radially varying (1D, one-dimensional) Earth model: the first value is lithosphere thickness (km), other values reflect mantle viscosity ($\times 10^{21}$ Pa s) for specific layers; see relevant publication.

^cGIA model details: SH = spherical harmonic (maximum degree indicated), FE = finite element, C = compressible, IC = incompressible, RF = rotational feedback, SG = self-gravitation, OL = ocean loading, 'x' = feature not included.

^dRSL = relative sea-level data; GPS rates corrected for elastic response to contemporary ice mass change.

^eEarth model taken from ref. ⁵¹.

^fIce model taken from ref. ⁵¹.

^gDifferent to ICE-6G_C in Antarctica, owing to the use of Bedmap2¹⁰⁷ topography.

Extended Data Table 2 | Details of the SMB models used in this study

Contributor	Model	Publication ^a	Class ^b	Area (10 ⁶ km ²)	Grid	SMB ^c (Gt/yr)	Precipitation ^c (Gt/yr)	Runoff ^c (Gt/yr)
Noël	RACMO2.3	Noël et al., 2015	RCM	1.73	11 km	350	721	311
Noël	RACMO2.3p2	Noël et al., 2018	RCM	1.73	11 km	432	727	258
Langen	HIRHAM5	Lucas-Picher et al., 2012	RCM	1.71	5.5 km	385	794	351
Fettweis	MARv3.6	Fettweis et al., 2017	RCM	1.69	7.5 km	381	706	308
Noël	RACMO2.3d	Noël et al., 2016	RCM-d	1.69	1 km	314	755	397
Noël	RACMO2.3p2d	Noël et al., 2018	RCM-d	1.69	1 km	338	703	331
Cullather	MERRA-2	Gelaro et al., 2017	GA-n	1.73	0.5 °	504	818	277
Hanna	ECMWF	Hanna et al., 2012	GA-d	1.65	5 km	370	532	186
Wilton	ECMWFd	Wilton et al., 2017	GA-d	1.71	1 km	314	603	246
Mernild	Snow Model	Mernild et al., 2010	PM	1.64	5 km	125	655	418

Refs. ^{9,13,21,46,115–119}.

^aMain reference publication; additional references are provided in Supplementary Table 1.

^bSMB model class; regional climate model (RCM), global numerical analysis (GA), process model (PM). Native resolution (n) and downscaled (d) models are also identified.

^cAverages over the period 1980–2012 for the Greenland Ice Sheet excluding peripheral ice caps and using the drainage basins from ref. ³⁷.

Extended Data Table 3 | Rate of Greenland Ice Sheet mass change for 2005–2015

Technique	Mass balance (Gt/yr)	s.d.(Gt/yr)	range (Gt/yr)
Altimetry*	-244 ± 15	43	122
Gravimetry	-248 ± 18	31	104
Input-Output Method	-281 ± 25	23	48
All	-255 ± 20	41	163

Estimates of ice-sheet mass balance from satellite altimetry, gravimetry the input–output method and from all three groups during the period 2005–2015. The average s.d. and ranges of individual estimates within each group during the same period are also shown. *No altimetry data in 2010.

Ruthenium isotope vestige of Earth's pre-late-veneer mantle preserved in Archaean rocks

<https://doi.org/10.1038/s41586-020-2069-3>

Received: 14 August 2019

Accepted: 15 January 2020

Published online: 11 March 2020

 Check for updates

Mario Fischer-Gödde^{1✉}, Bo-Magnus Elfers¹, Carsten Münker¹, Kristoffer Szilas², Wolfgang D. Maier³, Nils Messling¹, Tomoaki Morishita^{4,5,6}, Martin Van Kranendonk⁷ & Hugh Smithies⁸

The accretion of volatile-rich material from the outer Solar System represents a crucial prerequisite for Earth to develop oceans and become a habitable planet^{1–4}. However, the timing of this accretion remains controversial^{5–8}. It has been proposed that volatile elements were added to Earth by the late accretion of a late veneer consisting of carbonaceous-chondrite-like material after core formation had ceased^{6,9,10}. This view could not be reconciled with the ruthenium (Ru) isotope composition of carbonaceous chondrites^{5,11}, which is distinct from that of the modern mantle¹², or of any known meteorite group⁵. As a possible solution, Earth's pre-late-veneer mantle could already have contained a fraction of Ru that was not fully extracted by core formation¹³. The presence of such pre-late-veneer Ru can only be established if its isotope composition is distinct from that of the modern mantle. Here we report the first high-precision, mass-independent Ru isotope compositions for Eoarchaeon ultramafic rocks from southwest Greenland, which display a relative ¹⁰⁰Ru excess of 22 parts per million compared with the modern mantle value. This ¹⁰⁰Ru excess indicates that the source of the Eoarchaeon rocks already contained a substantial fraction of Ru before the accretion of the late veneer. By 3.7 billion years ago, the mantle beneath southwest Greenland had not yet fully equilibrated with late accreted material. Otherwise, no Ru isotopic difference relative to the modern mantle would be observed. If constraints from other highly siderophile elements besides Ru are also considered¹⁴, the composition of the modern mantle can only be reconciled if the late veneer contained substantial amounts of carbonaceous-chondrite-like materials with their characteristic ¹⁰⁰Ru deficits. These data therefore relax previous constraints on the late veneer and are consistent with volatile-rich material from the outer Solar System being delivered to Earth during late accretion.

Ruthenium is a highly siderophile element (HSE) and is therefore expected to be sequestered in the metallic core during Earth's differentiation. Contrary to this prediction, the abundances of Ru and other HSEs in the modern mantle are higher than expected compared with metal–silicate equilibrium conditions^{15,16}. This observation is most commonly explained by HSE replenishment of the mantle through the addition of a late veneer after core formation. Relative abundances of HSEs that are close to chondritic compositions in the mantle suggest that the late veneer must have consisted of primitive meteoritic material^{17,18}, amounting to ~0.5% of Earth's mass¹⁸. The chemical composition of the late veneer and its origin are a longstanding matter of debate, especially in the context of how and when Earth accreted its water and

volatiles^{3,6,9,10}. Previous studies debated whether significant amounts of volatile-rich carbonaceous-chondrite-like material were added by the late veneer during the final stages of Earth's accretion^{6,9,10} or had already been incorporated during earlier stages of Earth's growth^{3,5,7,8,11}.

Mass-independent ruthenium isotopic variations among meteorites and Earth have provided evidence that the late veneer was derived from reduced and volatile-poor inner Solar System materials most similar to enstatite chondrites^{5,11,12,19}. This is in contrast to constraints from the relative abundances of volatile elements such as selenium (Se), tellurium (Te) and sulfur (S) and the Se isotope composition in the silicate Earth that were used to argue for a CM or CI carbonaceous-chondrite-like late veneer composition^{2,9,10}. Owing to its distinct Ru

¹Institut für Geologie und Mineralogie, University of Cologne, Cologne, Germany. ²Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark. ³School of Earth and Ocean Sciences, Cardiff University, Cardiff, UK. ⁴Faculty of Geosciences and Civil Engineering, Institute of Science and Engineering, Kanazawa University, Kanazawa, Japan. ⁵Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA. ⁶Volcanoes and Earth's Interior Research Center, Research Institute for Marine Geodynamics, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan. ⁷Australian Centre for Astrobiology, University of New South Wales, Sydney, New South Wales, Australia. ⁸Geological Survey of Western Australia, East Perth, Western Australia, Australia. ✉e-mail: mfgisch48@uni-koeln.de

isotope composition, volatile-rich carbonaceous-chondrite-like material from the outer Solar System was excluded as possible late-veneer source material^{5,11}, and thus the late veneer seemed unlikely to be the primary source of water and volatiles on Earth^{5,11}. It should be noted, however, that this conclusion depends on the premise that the Ru in Earth's mantle originates solely from the late accreted materials that were added after cessation of core formation^{11,15,16,18}. If Earth's pre-late-veneer mantle retained a significant fraction of Ru during metal–silicate differentiation^{13,20}, as recently suggested, this conclusion would be invalid. Investigating Ru isotope signatures in the putative remnants of pre-late-veneer mantle would thus not only provide insights into the timescales and efficiencies of mixing the late veneer into Earth's mantle, but also introduce constraints on the composition of the material that was added as a late veneer.

To our knowledge, no unambiguous isotopic evidence for the preservation of pre-late-veneer mantle on Earth existed until now. For instance, resolvable excesses in ^{182}W reported for 3.8-billion-year-old (Gyr-old) Archaean rocks from Isua (Greenland) and Acasta (Canada) in conjunction with relatively low HSE abundances observed in 3.5–3.2-Gyr-old Archaean komatiites from the Pilbara Craton (Australia) and the Barberton greenstone belt (South Africa) were interpreted to reflect sluggish mixing of the late veneer into the early Archaean mantle^{21,22}. However, it was later suggested that the mantle sources of the 3.8–3.7-Gyr-old Isua supracrustal belt (ISB) rocks, including 3.8-Gyr-old Eoarchaean peridotites from the Narssaq ultramafic body (NUB) and the south of the Isua supracrustal belt (SOISB), already had HSE abundances at about 60–100% of the modern mantle value^{14,23}. This suggests that the late veneer was to a large extent mixed into the ambient mantle by ~3.8 billion years ago (Ga). To reconcile ^{182}W excesses with the presence of modern-mantle-like HSE abundances, it was proposed that a small amount of core material could have been entrained into proto-Earth's mantle as a consequence of the Moon-forming giant impact^{20,24}. However, ^{182}W anomalies could also be generated by early mantle differentiation processes during approximately the first 50 Myr of the Solar System^{25–29} or by core–mantle interactions in the sources of mantle plumes³⁰. In summary, ^{182}W and HSE concentration data alone fail to provide an unambiguous test of whether pre-late-veneer mantle domains were preserved.

Here we explore the potential use of mass-independent Ru isotope variations in terrestrial rocks as a tool to investigate whether pre-late-veneer isotope signatures can be found in the Archaean mantle. While the Ru isotope composition of the modern mantle is well constrained¹², this is not the case for the Archaean mantle. To address this issue, we determined the Ru isotope composition for a set of ultramafic rocks from different Eoarchaean and Palaeoproterozoic terranes (Extended Data Table 1; see Methods for details). We focus on the $^{100}\text{Ru}/^{101}\text{Ru}$ and $^{102}\text{Ru}/^{101}\text{Ru}$ ratios to constrain the Ru isotope compositions of the mantle sources of these rocks because these isotope ratios are measured at the highest precision and also show the largest variability among meteoritic materials^{5,19,31}. The results are reported as ϵ unit (0.01%) deviations of mass bias-corrected $^{100}\text{Ru}/^{101}\text{Ru}$ and $^{102}\text{Ru}/^{101}\text{Ru}$ ratios from a terrestrial standard.

Exotic composition of Archaean mantle

We report Ru isotope data for samples from four different cratons. The Ru isotope compositions obtained for ultramafic samples from the Pilbara Craton (3.5–3.2 Gyr old), the Superior Province (Abitibi greenstone belt, 2.7 Gyr old) and the Kaapvaal Craton (Bushveld Complex, 2.05 Gyr old) are indistinguishable from the Ru solution standard (Fig. 1), indicating that their Ru isotope compositions reflect that of the modern terrestrial mantle. By contrast, Eoarchaean 3.8–3.7-Gyr-old ultramafic rocks from the North Atlantic Craton, originating from various localities of the Itsaq gneiss complex (IGC) in southwest Greenland (the NUB, SOISB, ISB and the Ujaragssuit Nunât layered

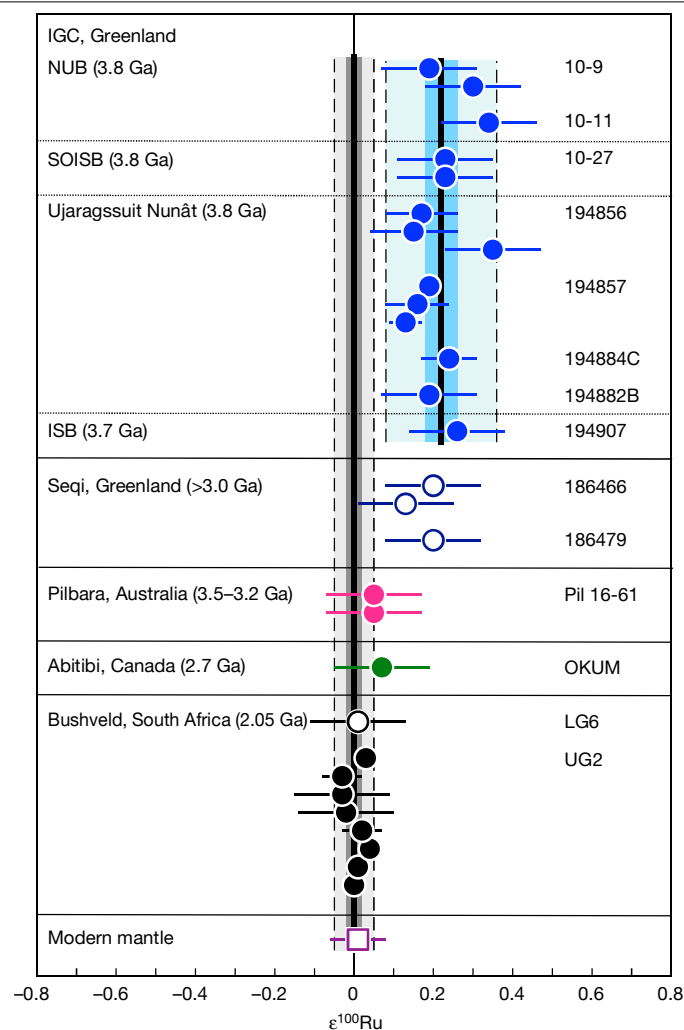


Fig. 1 | $\epsilon^{100}\text{Ru}$ data for Archaean and Palaeoproterozoic rocks, the modern mantle and chondrites. The individual results for all analysed samples (Extended Data Table 1) are shown with the composition of the modern mantle¹². The uncertainties for individual data points reflect the external uncertainty of the method (2 s.d. for samples measured $n < 4$ times) or 95% confidence intervals of replicate analyses of a given sample (if $n \geq 4$). The mean values for 3.8–3.7-Gyr-old Eoarchaean samples from the IGC in southwest Greenland and chondrite samples from the Bushveld complex are shown as solid vertical black lines. The darker grey and blue areas represent the respective 95% confidence intervals; the light grey and blue areas limited by dashed lines indicate the 2 s.d. uncertainty of the mean values. The uncertainty for the modern mantle composition is 2 s.d. (ref. 12). Numbers on the right of the data points refer to the sample identifiers given in Extended Data Table 1.

intrusion) exhibit a uniform and well-resolved excess in $\epsilon^{100}\text{Ru}$ of $+0.22 \pm 0.04$ (95% confidence interval, Fig. 1) combined with a smaller excess in $\epsilon^{102}\text{Ru}$ of $+0.09 \pm 0.02$ (95% confidence interval, Fig. 2a). Chondrites from the younger 3.0-Gyr-old Seqi ultramafic complex in southwest Greenland show the same excesses in $\epsilon^{100}\text{Ru}$ and $\epsilon^{102}\text{Ru}$. The combined $\epsilon^{100}\text{Ru}$ and $\epsilon^{102}\text{Ru}$ excesses in these rocks represent mass-independent isotope anomalies of nucleosynthetic origin and indicate that the Ru in the southwest Greenland mantle source is enriched in nuclides produced by the slow neutron capture process (s-process) of nucleosynthesis compared with the modern mantle (Fig. 2a). The isotope excesses cannot be explained by mass-independent fractionation effects or by inherited fissiogenic Ru nuclides (see Methods and Extended Data for details about the accuracy of the Ru isotope data).

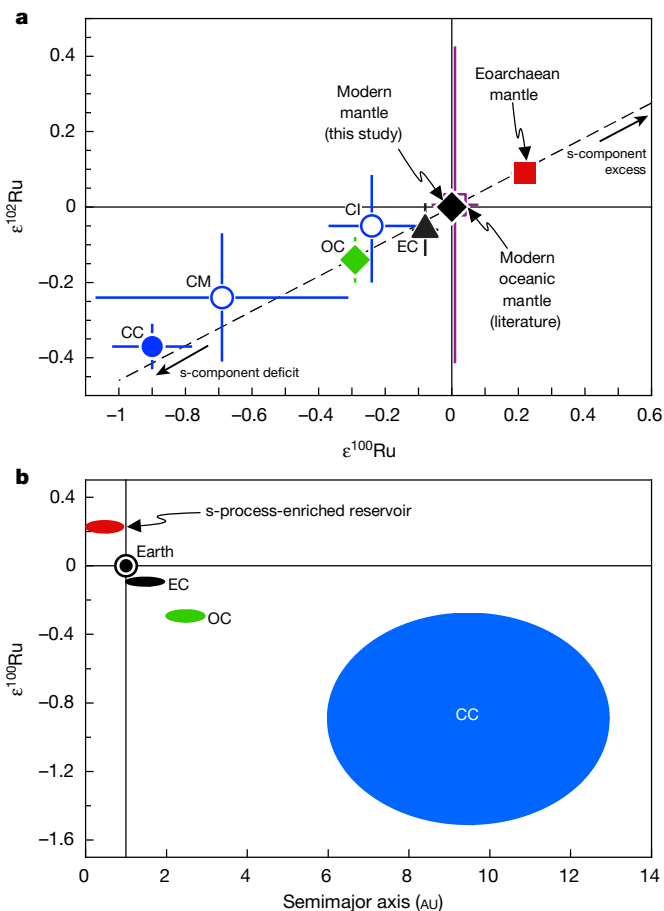


Fig. 2 | Ru isotope plot illustrating compositional differences between enstatite, ordinary, average carbonaceous, CI and CM carbonaceous chondrites, the modern mantle and the Eoarchaeon mantle. a, The dashed line represents a mixing line between the modern mantle composition ($\epsilon^{100}\text{Ru} = 0$) and an s-process component defined by Ru isotope data for pre-solar silicon carbide grains³⁴. The compositions of enstatite chondrites (EC, $\epsilon^{100}\text{Ru} = -0.08 \pm 0.04$, 95% confidence interval); ordinary chondrites (OC, $\epsilon^{100}\text{Ru} = -0.29 \pm 0.03$, 95% confidence interval)⁵, CI chondrites (CI, $\epsilon^{100}\text{Ru} = -0.24 \pm 0.13$, 2 s.d.)⁵; CM chondrites (CM, $\epsilon^{100}\text{Ru} = -0.69 \pm 0.38$, 95% confidence interval)⁵, and average carbonaceous chondrites (average CC, $\epsilon^{100}\text{Ru} = -0.90 \pm 0.12$, 95% confidence interval)⁵ are shown for comparison. The uncertainties for CI chondrites reflect a single measurement and are thus shown with the external uncertainty of the method (2 s.d. as stated in ref. ⁵). Uncertainties for the modern and the Eoarchaeon mantle composition are the same as stated in Fig. 1. Note that the uncertainty for the modern oceanic mantle composition from the literature is shown as 2 s.d. (ref. ¹²). **b,** Heliocentric zoning of $\epsilon^{100}\text{Ru}$ anomalies⁵. The presence of an s-process-enriched reservoir that contributed to Earth's growth is inferred from the Ru isotope composition obtained for the Eoarchaeon mantle of southwest Greenland (Fig. 1). Chondrite groups formed at increasing heliocentric distances exhibit more negative $\epsilon^{100}\text{Ru}$ because they are more depleted in s-process Ru relative to Earth's modern mantle⁵. The $\epsilon^{100}\text{Ru}$ uncertainty for carbonaceous chondrites in **b** is shown as 2 s.d. to account for the significant within-group variation of their $\epsilon^{100}\text{Ru}$ values (image adapted from ref. ⁵, Springer Nature).

The s-process-enriched composition inferred for the Archaean southwest Greenland mantle is an unexpected finding because the Ru isotope compositions reported for all meteorites are deficient in s-process Ru and exhibit negative $\epsilon^{100}\text{Ru}$ and $\epsilon^{102}\text{Ru}$ anomalies relative to Earth's modern mantle^{5,19,31}. The southwest Greenland data provide unambiguous evidence for s-process-enriched building material that contributed to the early stages of Earth's growth. Owing to the observed heliocentric zoning of $\epsilon^{100}\text{Ru}$ anomalies among meteorites⁵, we speculate that this

reservoir was most probably located in the innermost region of the Solar System, within 1 astronomical unit (Fig. 2b).

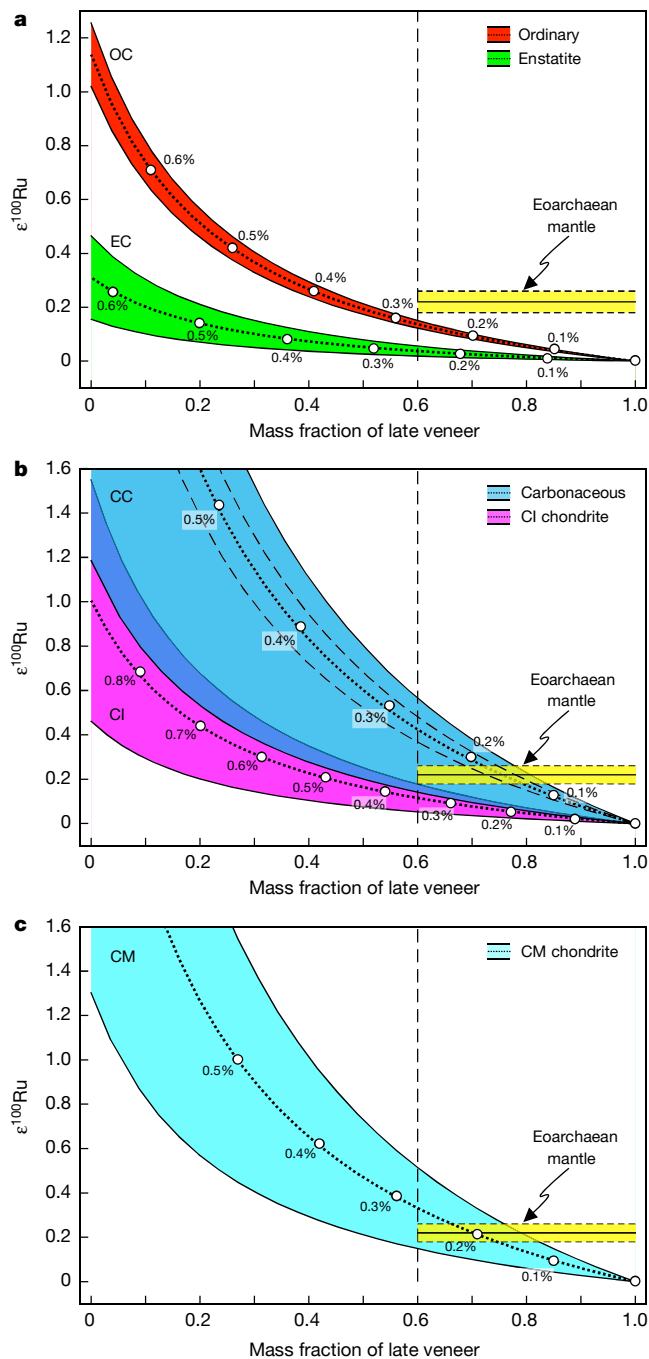
Pre-late-veener Ru isotopic signature

The ^{100}Ru excess provides unequivocal evidence that the mantle source of the Greenland rocks did not receive the full complement of late veneer material²¹. Furthermore, it also requires that Ru (and possibly other HSEs) was not completely stripped from the mantle during the latest stages of core formation¹³. Otherwise, no Ru isotope anomaly would be observed. The uniform and ubiquitous presence of the $\epsilon^{100}\text{Ru}$ anomaly in various 3.8–3.7-Gyr-old ultramafic rock types from different Eoarchaeon terranes in Greenland (Isuakasia, Færingehaven) suggests that a larger mantle domain is lacking a full late veneer component^{20,21}. The presence of the $\epsilon^{100}\text{Ru}$ anomaly in the younger Mesarchaeon chromitites from Seqi (minimum age of 3.0 Gyr, Akia terrane) also indicates that even 700 Myr later, the southwest Greenland mantle had not fully equilibrated with the late veneer. Such a prolonged timescale for mixing-in of the late veneer component is consistent with significant HSE depletions observed in Archaean mafic rocks from the Pilbara and Kaapvaal cratons²², which previously had been explained by sluggish inmixing of late veneer material.

As outlined above, the $\epsilon^{100}\text{Ru}$ excess identified in Eoarchaeon ultramafic rocks from southwest Greenland indicates that Ru was not completely sequestered in the core, most probably because some late accretionary component had been delivered during the waning stages of core formation^{14,23,32}. Depending on the composition of this early late veneer material, the ^{100}Ru excess measured in the Greenland rocks would then represent a minimum estimate for the $\epsilon^{100}\text{Ru}$ excess of the pure pre-late-veener mantle. The nature of the early component that supplied the ^{100}Ru excess and was already mixed into the Greenland mantle before 3.8 Ga, probably inner Solar System material (Fig. 2b), can be further constrained by osmium isotope systematics. This is because the initial osmium isotopic compositions of chromitite and peridotite samples from the IGC overlap the $^{187}\text{Os}/^{188}\text{Os}$ composition of chondrites at 3.8 Ga (refs. ^{14,23,32,33}) (Extended Data Table 2). Assuming that the positive $\epsilon^{100}\text{Ru}$ anomaly and the chondritic Os signature were both imparted by this component, it is unlikely that it is represented by any known chondritic meteorites because these all exhibit negative $\epsilon^{100}\text{Ru}$ values and chondritic Os isotope compositions. Importantly, owing to its positive $\epsilon^{100}\text{Ru}$ value, this material cannot derive from a carbonaceous-chondrite-like Moon-forming impactor⁸ because carbonaceous chondrites also exhibit the most negative $\epsilon^{100}\text{Ru}$ values among all known chondrite groups⁵ (Fig. 2).

Carbonaceous-chondrite-like late veneer

Regardless of the precise nature and origin of the early accreted component, the ^{100}Ru excess inferred for the Eoarchaeon southwest Greenland mantle source could only be balanced by the addition of chondritic materials with negative $\epsilon^{100}\text{Ru}$ to yield the composition of the modern mantle ($\epsilon^{100}\text{Ru} = 0$). This mixing relationship is further illustrated in Fig. 3, where possible $\epsilon^{100}\text{Ru}$ compositions for the pre-late-veener mantle are calculated by subtracting enstatite, ordinary or carbonaceous-chondrite-like materials from the Ru isotopic composition of the modern mantle. The model is based on a recently proposed inefficient core formation scenario where about 20% of the Ru ($\sim 1.4 \text{ ng g}^{-1}$) in the modern mantle derives from the pre-late-veener stage¹³. Assuming a minimum late accretion contribution of 60% for the ≥ 3.8 -Gyr-old Itsaq mantle source¹⁴, only the addition of a late veneer consisting of carbonaceous chondrites could account for $\epsilon^{100}\text{Ru} \approx 0$, as observed for the modern mantle¹² (Fig. 3). The required proportion of late accreted material would amount to a maximum estimate of 0.3% of Earth's mass of average carbonaceous chondrite or CM carbonaceous chondrite material, consistent with a recent estimate based on Se isotopes¹⁰.



A late veneer consisting of CM-like material is also supported by the abundances of volatile elements in the silicate Earth^{2,9,10}. This conclusion remains robust, even if a larger fraction of Ru was present in the pre-late-veener mantle or the late accretion component in the mantle by 3.8 Ga was >60%, but in these cases a lower carbonaceous chondrite mass fraction would be sufficient. Ordinary chondrites would only become viable late veneer materials if the Greenland mantle contained a significantly lower late veneer contribution by 3.8 Ga (<50%). A late veneer consisting of carbonaceous chondrites is consistent with the relative abundances of S–Se–Te and the Se isotopic composition of the modern mantle^{2,9,10}, but the addition of a late veneer composed of ordinary chondrites cannot be reconciled with these constraints, because the relative abundances of S–Se–Te and the Se isotope composition of ordinary chondrites are distinct from those of Earth's mantle^{9,10}. If a major part of the late veneer consisted of core fragments from differentiated impactors²⁴, one potential caveat would be that this material

Fig. 3 | Model estimates for the amount of late veneer added to the Eoarchaeon mantle based on Ru isotope compositions. **a–c**, The model illustrates the effect of subtracting variable mass fractions of late veneer from the Ru isotope composition of the modern mantle. The composition of the modern mantle was fixed at $\epsilon^{100}\text{Ru} = 0$, as indicated by the composition obtained for samples from the Bushveld complex (Extended Data Table 1), which is indistinguishable from previously reported data for the modern oceanic mantle composition¹². Dotted lines indicate the respective mass fractions of different chondritic late-veener compositions subtracted from the modern mantle composition. Solid black lines show the minimum and maximum $\epsilon^{100}\text{Ru}$ values for different chondrite classes in **a**: enstatite chondrites (EC, $\epsilon^{100}\text{Ru} = -0.08 \pm 0.04$, 95% confidence interval) and ordinary chondrites (OC, $\epsilon^{100}\text{Ru} = -0.29 \pm 0.03$, 95% confidence interval)⁵; **b**: carbonaceous chondrites (CC, $\epsilon^{100}\text{Ru} = -0.90 \pm 0.61$, 2 s.d.; ± 0.12 , 95% confidence interval)⁵ and CI chondrites (CI, $\epsilon^{100}\text{Ru} = -0.24 \pm 0.13$, 2 s.d.)⁵; and **c**: CM chondrites (CM, $\epsilon^{100}\text{Ru} = -0.69 \pm 0.38$, 95% confidence interval)⁵. The uncertainties for carbonaceous chondrites in **b** are given as 2 s.d. to account for significant within-group variation in their $\epsilon^{100}\text{Ru}$ values (the dashed line indicates 95% confidence interval uncertainty). The 2 s.d. uncertainty for CI chondrites reflects the external uncertainty of the method (as stated in ref. 5). The amount of subtracted late veneer material for each respective chondrite composition was adjusted to match a presumed Ru concentration in the pre-late-veener mantle of -1.4 ng g^{-1} , corresponding to -20% of the Ru contained in the modern mantle. The yellow boxes indicate the composition of the Eoarchaeon mantle inferred from the mean value of all analysed 3.8–3.7-Gyr-old samples originating from various localities of the IGC ($\epsilon^{100}\text{Ru} = +0.22 \pm 0.04$, average value shown as solid black line, dashed lines indicate 95% confidence interval uncertainty, Extended Data Table 1). The solid vertical dashed line indicates the minimum late veneer contribution inferred for the mantle source of peridotite samples from the NUB and the SOISB based on previously reported $^{187}\text{Os}/^{188}\text{Os}$ data and HSE concentrations¹⁴. The parameters used for the mixing model are given in Extended Data Table 3.

cannot readily account for chondritic S–Se–Te and broadly chondritic HSE relative abundances in Earth's mantle^{2,9,16,18}.

Collectively, our data imply that the distinct ^{100}Ru isotope excess in the Eoarchaeon southwest Greenland mantle source is best explained by late mixing of a carbonaceous-chondrite-like late veneer fraction into Earth's mantle. Thus, contrary to previous Ru isotope constraints on the late veneer^{5,11}, these data imply that significant amounts of volatile-rich outer Solar System materials including water and volatiles were added with the late veneer. This revised view also agrees with other constraints, such as those independently obtained from the relative abundances and isotope compositions of Earth's volatile elements^{1,2,9,10}, which also indicate that the major share of Earth's volatiles was inherited from a carbonaceous chondrite source^{1–4}. Finally, our data demonstrate that investigating the Ru isotope composition of terrestrial rocks represents a powerful analytical tool for identifying primordial mantle heterogeneities arising from incomplete equilibration of the ambient mantle with Earth's late-stage building blocks.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2069-3>.

- Alexander, C. M. et al. The provenances of asteroids, and their contributions to the volatile inventories of the terrestrial planets. *Science* **337**, 721–723 (2012).
- Braukmüller, N., Wombacher, F., Funk, C. & Münker, C. Earth's volatile element depletion pattern inherited from a carbonaceous chondrite-like source. *Nat. Geosci.* **12**, 564–568 (2019).
- Marty, B. The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth Planet. Sci. Lett.* **313–314**, 56–66 (2012).

4. Peslier, A. H., Schönbachler, M., Busemann, H. & Karato, S.-I. Water in the Earth's interior: distribution and origin. *Space Sci. Rev.* **212**, 743–810 (2017).
5. Fischer-Gödde, M. & Kleine, T. Ruthenium isotopic evidence for an inner Solar System origin of the late veneer. *Nature* **541**, 525–527 (2017).
6. Albarède, F. Volatile accretion history of the terrestrial planets and dynamic implications. *Nature* **461**, 1227–1233 (2009).
7. Schönbachler, M., Carlson, R. W., Horan, M. F., Mock, T. D. & Hauri, E. H. Heterogeneous accretion and the moderately volatile element budget of Earth. *Science* **328**, 884–887 (2010).
8. Budde, G., Burkhardt, C. & Kleine, T. Molybdenum isotopic evidence for the late accretion of outer Solar System material to Earth. *Nat. Astron.* **3**, 736–741 (2019).
9. Wang, Z. & Becker, H. Ratios of S, Se and Te in the silicate Earth require a volatile-rich late veneer. *Nature* **499**, 328–331 (2013).
10. Varas-Reus, M. I., König, S., Yierpan, A., Lorand, J. P. & Schoenberg, R. Selenium isotopes as tracers of a late volatile contribution to Earth from the outer Solar System. *Nat. Geosci.* **12**, 779–782 (2019).
11. Dauphas, N. The isotopic nature of the Earth's accreting material through time. *Nature* **541**, 521–524 (2017).
12. Bermingham, K. R. & Walker, R. J. The ruthenium isotopic composition of the oceanic mantle. *Earth Planet. Sci. Lett.* **474**, 466–473 (2017).
13. Rubie, D. C. et al. Highly siderophile elements were stripped from Earth's mantle by iron sulfide segregation. *Science* **353**, 1141–1144 (2016).
14. van de Löcht, J. et al. Earth's oldest mantle peridotites show entire record of late accretion. *Geology* **46**, 199–202 (2018).
15. Brenan, J. M. & McDonough, W. F. Core formation and metal-silicate fractionation of osmium and iridium from gold. *Nat. Geosci.* **2**, 798–801 (2009).
16. Becker, H. et al. Highly siderophile element composition of the Earth's primitive upper mantle: Constraints from new data on peridotite massifs and xenoliths. *Geochim. Cosmochim. Acta* **70**, 4528–4550 (2006).
17. Chou, C. L. Fractionation of siderophile elements in the Earth's upper mantle. *Proc. Lunar Planet. Sci. Conf.* **9**, 219–230 (1978).
18. Walker, R. J. Highly siderophile elements in the Earth, Moon and Mars: update and implications for planetary accretion and differentiation. *Geochemistry* **69**, 101–125 (2009).
19. Bermingham, K. R., Worsham, E. A. & Walker, R. J. New insights into Mo and Ru isotope variation in the nebula and terrestrial planet accretionary genetics. *Earth Planet. Sci. Lett.* **487**, 221–229 (2018).
20. Willbold, M., Mojzsis, S. J., Chen, H. W. & Elliott, T. Tungsten isotope composition of the Acasta Gneiss Complex. *Earth Planet. Sci. Lett.* **419**, 168–177 (2015).
21. Willbold, M., Elliott, T. & Moorbath, S. The tungsten isotopic composition of the Earth's mantle before the terminal bombardment. *Nature* **477**, 195–198 (2011).
22. Maier, W. D. et al. Progressive mixing of meteoritic veneer into the early Earth's deep mantle. *Nature* **460**, 620–623 (2009).
23. Dale, C. W., Kruijer, T. S. & Burton, K. W. Highly siderophile element and ^{182}W evidence for a partial late veneer in the source of 3.8 Ga rocks from Isua, Greenland. *Earth Planet. Sci. Lett.* **458**, 394–404 (2017).
24. Marchi, S., Canup, R. M. & Walker, R. J. Heterogeneous delivery of silicate and metal to the Earth by large planetesimals. *Nat. Geosci.* **11**, 77–81 (2018).
25. Touboul, M., Puchtel, I. S. & Walker, R. J. ^{182}W evidence for long-term preservation of early mantle differentiation products. *Science* **335**, 1065–1069 (2012).
26. Rizo, H. et al. Preservation of Earth-forming events in the tungsten isotopic composition of modern flood basalts. *Science* **352**, 809–812 (2016).
27. Mundl, A. et al. Tungsten-182 heterogeneity in modern ocean island basalts. *Science* **356**, 66–69 (2017).
28. Touboul, M., Liu, J., O'Neil, J., Puchtel, I. S. & Walker, R. J. New insights into the Hadean mantle revealed by ^{182}W and highly siderophile element abundances of supracrustal rocks from the Nuvvuagittuq Greenstone Belt, Quebec, Canada. *Chem. Geol.* **383**, 63–75 (2014).
29. Tusch, J. et al. Uniform ^{182}W isotope compositions in Eoarchean rocks from the Isua region, SW Greenland: the role of early silicate differentiation and missing late veneer. *Geochim. Cosmochim. Acta* **257**, 284–310 (2019).
30. Rizo, H. et al. ^{182}W evidence for core-mantle interaction in the source of mantle plumes. *Geochim. Perspect. Lett.* **11**, 6–11 (2019).
31. Chen, J. H., Papanastassiou, D. A. & Wasserburg, G. J. Ruthenium endemic isotope effects in chondrites and differentiated meteorites. *Geochim. Cosmochim. Acta* **74**, 3851–3862 (2010).
32. Szilas, K., Kelemen, P. B. & Rosing, M. T. The petrogenesis of ultramafic rocks in the >3.7 Ga Isua supracrustal belt, southern West Greenland: geochemical evidence for two distinct magmatic cumulate trends. *Gondwana Res.* **28**, 565–580 (2015).
33. Bennett, V. C., Nutman, A. P. & Esat, T. M. Constraints on mantle evolution from $^{187}\text{Os}/^{188}\text{Os}$ isotopic compositions of Archean ultramafic rocks from southern West Greenland (3.8 Ga) and Western Australia (3.46 Ga). *Geochim. Cosmochim. Acta* **66**, 2615–2630 (2002).
34. Savina, M. R. et al. Extinct technetium in silicon carbide stardust grains: implications for stellar nucleosynthesis. *Science* **303**, 649–652 (2004).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Samples

The samples analysed in this study comprise ultramafic rocks from four different cratons. The North Atlantic Craton is represented by peridotite and chromitite samples of from various localities of the Eoarchaeon IGC and the Mesoarchaeon Seqi ultramafic complex in southwest Greenland. The sample set is complemented by chromitites from the Kaapvaal Craton (Bushveld Complex, South Africa) and the Pilbara Craton (Australia) and a komatiite reference sample from the Superior Province (Abitibi greenstone belt, Canada). Where available, the Ru concentration and osmium isotope data for the samples analysed in this study are given in Extended Data Table 2. There were no data available for some samples, so Ru and Os data obtained for similar samples from the same location are listed.

The IGC in southwest Greenland represents one of the few localities where remnants of Eoarchaeon mantle are preserved. The IGC comprises two Eoarchaeon crustal terranes (Isuakasia and Færingehavn), where possible mantle rocks are exposed as ultramafic lenses in the 3.8–3.7-Gyr-old Isua supracrustal belt and ultramafic bodies in the 3.8-Gyr-old SOISB, both located in the Isuakasia terrane^{35,36}. In the Færingehavn terrane, such rocks are exposed in the 3.8-Gyr-old NUB³⁷. The peridotite and chromitite samples investigated in this study were selected to cover all of these different localities.

Samples 10-9 and 10-11 from the NUB are massive, coarse-grained peridotites. Olivine is the dominant phase in these rocks. They also contain orthopyroxene, amphibole, spinel and magnetite¹⁴. The chemical compositions of these samples, including concentration data for highly siderophile elements and ¹⁸⁷Os/¹⁸⁸Os data, were reported in a previous study¹⁴. Sample 10-27 is a harzburgite from the SOISB. The mineral assemblage of this rock is comparable to other harzburgites collected from the same locality¹⁴. These rocks are typically spinel-peridotites with harzburgitic mineral assemblages composed mainly of olivine and variable amounts of orthopyroxene, amphibole and opaque phases¹⁴. A minimum age of 3.8 Gyr was estimated for the analysed SOISB and NUB peridotites on the basis of field relationships with surrounding 3.8-Gyr-old tonalitic gneisses and crosscutting dykes^{14,36,37}.

Two of the investigated chromitites (194856, 194857) were collected from the Ujargssuit Nunât layered intrusion³⁸. For chromitites from this locality, Pt–Os model ages as old as 4.36 Gyr were reported³⁹. Samples 194882B and 19488C are chromitites from a locality close to the inland ice that most probably belongs to the same sequence as the chromitites from the Ujargssuit Nunât layered intrusion.

The dunite sample 194907 was collected from an antigorite lens located within the northeastern part of the 3.7-Gyr-old Isua supracrustal belt, which has previously been referred to as Dunite Lens B^{40,41}. The ISB dunites also contain orthopyroxene and spinel, and very minor amounts of clinopyroxene that has mostly been altered⁴⁰.

Two of the analysed chromitites (186466, 186479) derive from the Seqi ultramafic complex. The major and trace element compositions, including concentration data for platinum group elements of these samples, were reported in a previous study⁴². The Seqi ultramafic complex represents a peridotite enclave hosted by tonalitic orthogneiss within the 3.0-Gyr-old Akia terrane. A minimum age for the ultramafic body is constrained by 2.98-Gyr-old crosscutting granitoid sheets⁴², although unpublished Re–Os isotope data show a consistent 3.1 Gyr mantle depletion age for the Seqi ultramafic complex. The highly refractory peridotites and chromitites are interpreted as representing the remnant of a fragmented layered complex or a magma conduit. The ultramafic rocks formed from a magnesian-rich, near-anhydrous magma as olivine dominated cumulates with high modal contents of chromite⁴². Their parental magma was generated by high degrees of partial melting of a mantle source that probably represents the precursor of the regional sub-continental lithospheric mantle.

The chromitite sample Pil 16-61 was collected from the Warawoona Group located within the Pilbara Craton. The chromitite may be as old as the associated Mount Ada basalt unit (3.5 Gyr) or it may be part of the younger 3.2-Gyr-old Dalton Suite sill complex⁴³.

The investigated komatiite rock (OKUM) from the 2.7-Gyr-old Abitibi greenstone belt (Canada) is a commercially available rock reference sample provided by the International Association of Geoanalysts.

Two chromitites (UG2, LG6) from the 2.05-Gyr-old Bushveld complex (South Africa) were used as a reference sample to validate the analytical method and to assess the precision and the accuracy of the Ru isotope measurements. The Ru isotope composition of UG-2 was previously determined employing a different digestion method (alkaline fusion)⁵.

Ruthenium separation and purification

The required amount of sample material to yield sufficient Ru for a high-precision measurement was estimated on the basis of previously reported Ru concentrations (10-9, 10-11, UG-2, LG-6, OKUM, 186466, 186479)^{42,44,45} or Ru concentrations reported for samples of similar composition from the same locality (10-27, 194907)^{23,46}. When information was not available (for example, for samples 194856, 194857 and Pil16-61), the Ru concentrations were determined from a 1 g powder test portion digested in a high-pressure asher in reverse aqua regia (5 ml concentrated HNO₃ and 2.5 ml concentrated HCl). Before quantification of Ru for these samples, the digestion solution was dried down, converted twice with 5 ml of 6 M HCl, taken up in 0.2 M HCl and loaded on a cation column to remove matrix elements as described below. Ruthenium concentrations were determined in the eluted Ru fractions by external calibration using a quadrupole inductively coupled plasma mass spectrometer (ThermoScientific iCap). We note that the concentrations determined by this procedure may underestimate the actual concentration of samples because some Ru may have been lost as a volatile tetroxide (RuO₄) when the aqua regia solutions were dried down. These concentrations are therefore considered to be only approximate values. In a similar manner, the Ru contents of two chromitites (194882B, 194884C) were estimated from a 1% sample aliquot taken after NiS digestion and cation column chemistry as described below. However, we note that these estimates represent only approximate values too because the Ru yield of the NiS procedure is <100%.

For the NiS procedure, powder aliquots of 5–10 g were digested using a NiS fire assay technique⁴⁷. For chromitite samples with high Ru concentrations (UG-2, LG-6, 194856, 194857, Pil16-61), one NiS digestion with 5 g of sample powder was needed to yield sufficient amounts of Ru. Multiple NiS digestions with 10 g of sample powder had to be prepared for ultramafic samples with lower Ru concentrations (harzburgites, dunites, komatiites and some chromitites). The total number of NiS digestions and the amount of sample material used for each respective NiS bead are given in Extended Data Table 2. Appropriate amounts of Ni, S, borax and Na₂CO₃ were added to each 5–10 g sample portion and thoroughly mixed. The mixture was fluxed in a muffle furnace for 75 min at 1,000 °C. After cooling, the NiS beads were physically removed from the quenched silicate melt.

For the majority of samples, the NiS procedure resulted in about one to three beads of about 1 cm in diameter that could readily be recovered from the quenched silicate. The Ru yield of the NiS procedure was determined on the basis of sample powders with known Ru concentrations (UG-2, OKUM, 10-9, 10-11, 186466, 186479). The Ru yield for these samples usually varied from 60–95%. However, in case of three replicate UG-2 digestions the Ru yields were only of the order of 10–20%. The lower yields resulted from incomplete homogenization and subsequent inefficient extraction of NiS beads from the quenched silicate. The NiS digestions for these samples produced finely dispersed millimetre- to micrometre-sized spherules within the quenched silicate. Careful homogenization of the NiS sample–flux mixtures before digestion helped to avoid this problem.

The NiS beads were crushed in an agate mortar and transferred into 60 ml Savillex beakers to which 30 ml of concentrated HCl was added. The solutions were evaporated to near dryness on a hotplate at 100 °C. This step was repeated with another 30 ml of concentrated HCl and 20 ml of 1 M HCl.

Ruthenium was separated from the dissolved NiS beads using cation exchange chromatography⁴⁸. Each dissolved NiS bead from a single fire assay digestion was split over three cation columns filled with 10 ml AG50 X8 resin, respectively. The resin was equilibrated with 20 ml of 0.2 M HCl. Ruthenium and other platinum group elements were loaded and eluted in 14 ml of 0.2 M HCl. The eluted Ru fractions from each sample were recombined and a small aliquot (1%) was taken to determine the amount of Ru and remaining matrix elements (mainly Ni). If significant amounts of matrix elements passed through the column (if Ni/Ru > 1), the combined fractions of samples were passed for a second time over a single 10 ml cation column. The Ru yields from the cation column were usually >95%. The eluted sample solutions were dried down on a hotplate, recombined and Ru was further purified using a macrodistillation unit as described elsewhere⁴⁸. After the distillation, the purified Ru fractions were dried down on a hotplate and dissolved in 0.5 ml of 0.28 M HNO₃, from which a small aliquot was prepared as a predilution to determine the Ru yield and to check for potential interfering elements. The distillation yields were usually between 40 and 80%. The total Ru yield of the analytical procedure, including NiS digestion, column chemistry and distillation, is typically 30–70%, estimated from samples with known Ru concentrations (UG-2, OKUM, 10-9, 10-11, 186466, 186479). The total yield of the three UG-2 digestions was only 6–21%. The yields of the distillation for these samples were 50–80%, so the low total Ru yields are caused by inefficient extraction of NiS beads, as described above. However, neither the total Ru yield of the entire analytical procedure nor the respective yields from the NiS digestion or the Ru distillation have any effect on the accuracy of the Ru isotope data (Extended Data Fig. 1).

The procedural blank for a single NiS digestion, including column chemistry and distillation, varied between 185 and 435 pg ($n = 3$). The blank contribution was <1% for the majority of samples and <2% for OKUM and 194907 given that ≥ 30 ng of Ru were processed for each respective NiS digestion.

Mass spectrometry

The Ru isotope measurements were performed using a ThermoScientific Neptune Plus multicollector inductively coupled plasma mass spectrometer in the Institut für Geologie und Mineralogie at the University of Cologne. For the measurements, the Ru fractions were further diluted in 0.28 M HNO₃ to yield Ru solutions of 100 ng ml⁻¹. The diluted solutions were checked for the presence of interfering elements (Zr, Ni) that could affect the accuracy of the isotope data and cannot be monitored online during the measurements. The sample solutions were introduced into the mass spectrometer at an uptake rate of ~ 50 μ l min⁻¹ using an ESI microflow PFA nebulizer attached to a Cetac AridusII desolvator. The isotope measurements were conducted with total ion beam intensities between 8×10^{-11} and 2×10^{-10} A, obtained for 100 ng ml⁻¹ Ru sample and standard solutions using conventional NiH-cones. The set-up was optimized to yield oxide production rates <1% (CeO/Ce). The measurements were conducted in static mode and the seven stable Ru isotopes (⁹⁶Ru, ⁹⁸Ru, ⁹⁹Ru, ¹⁰⁰Ru, ¹⁰¹Ru, ¹⁰²Ru and ¹⁰⁴Ru) as well as ⁹⁷Mo and ¹⁰⁵Pd were monitored simultaneously. Each Ru isotope analysis consisted of an on-peak baseline on a solution blank (40 integrations of 4.2 s) followed by 100 integrations of 8.4 s for each sample or standard solution and typically consumed about 90 ng Ru. Each sample analysis was bracketed by measurements of an in-house Ru standard solution (Alfa Aesar Ru). The data were internally normalized to ⁹⁹Ru/¹⁰¹Ru = 0.7450754 (ref. ³¹) using the exponential law to corrected for mass-dependent isotope fractionation. The isotope data are reported as $\epsilon^i\text{Ru} = ([(^i\text{Ru}/^{101}\text{Ru})_{\text{Sample}} / (^i\text{Ru}/^{101}\text{Ru})_{\text{Standard}}] - 1) \times 1$

⁰⁴, calculated relative to the bracketing standard of each analytical session. The accuracy and precision of the Ru isotopic measurements were evaluated by replicate digestions and multiple analyses of the UG-2 chromitite (Bushveld), which was used as a reference sample. The Ru isotope data obtained for UG-2 in this study agree well with previously reported data⁵, where a different digestion method (alkaline fusion) was used for sample decomposition. This demonstrates that the isotope data obtained by the NiS method yield accurate results. The external reproducibilities (2 s.d.) obtained for a total number of 103 individual measurements from 8 replicate UG-2 digestions are ± 0.43 $\epsilon^{96}\text{Ru}$, ± 0.49 $\epsilon^{98}\text{Ru}$, ± 0.12 $\epsilon^{100}\text{Ru}$, ± 0.16 $\epsilon^{102}\text{Ru}$ and ± 0.30 $\epsilon^{104}\text{Ru}$.

Correction for mass-dependent isotope fractionation

The exponential law is one of the most commonly used methods to correct for natural and instrumental mass-dependent isotope fractionation. One potential caveat in using this correction for Ru isotope measurements could be that the distillation technique used to purify the Ru could induce an isotope fraction that would not follow the exponential law. This could cause apparent isotopic anomalies for a given sample as a consequence of inaccurate mass fractionation correction. The exponential law assumes that the logarithmic fractionation $\beta = \ln(r/r')$ of a given isotopic ratio is expressed as a function of the mass log difference $\Delta(\ln M) = \ln(M_2/M_1)$. Considering two isotopic ratios ($r = ^{99}\text{Ru}/^{101}\text{Ru}$ and $r' = ^{100}\text{Ru}/^{101}\text{Ru}$) the exponential law predicts that mass fractionation produces a linear array in a $\ln(r/r')$ plot⁴⁹. This is illustrated in Extended Data Fig. 2 for the measured raw ratios of ⁹⁹Ru/¹⁰¹Ru and ¹⁰⁰Ru/¹⁰¹Ru. The ratios in this figure are not corrected for mass fractionation and are normalized to a reference ratio (R and R' , respectively)³¹. If the mass fractionation experienced by the samples is accurately described by the exponential law, the ratios should fall on a linear array with a slope of -0.5. Two distinct mass fractionation lines can be observed in the plot for different sessions. The slopes for both groups of sessions are indistinguishable within error and are in very good agreement with the slope predicted by the exponential law. Most importantly, the samples purified by distillation fall on the same respective mass fractionation line as their associated Alfa Aesar bracketing standards. This clearly demonstrates that the Ru distillation does not induce any non-exponential mass-fractionation effects for the samples in comparison with the bracketing standard. This observation is also independent from the Ru yield of the samples and does not change if other Ru isotope ratios are considered. Thus, the Ru isotope anomalies obtained for the southwest Greenland samples cannot reflect inaccurate mass fractionation correction. The shift observed for samples and associated standards plotting on a distinct mass fractionation array in Extended Data Figure 2 was caused by maintenance in May 2019 during which a Faraday cup was replaced. However, because the data are reported as relative deviations in parts per 10⁴ from the Alfa Aesar bracketing standard, and because samples and bracketing standards are shifted by the same magnitude, this does not affect the accuracy of the data. This is also confirmed by replicate digestions of sample 10-9 that were analysed in both groups of sessions. The $\epsilon^{100}\text{Ru}$ values for this sample are indistinguishable within analytical uncertainty.

Another argument against non-exponential mass-fractionation effects is that nucleosynthetic Ru isotope anomalies caused by variable contributions of s-process Ru nuclides would not lead to any $\epsilon^{104}\text{Ru}$ anomalies in the ⁹⁹Ru/¹⁰¹Ru normalization scheme. As the $\epsilon^{104}\text{Ru}$ values for all analysed samples fall within the external reproducibility of the method (± 0.30 for $\epsilon^{104}\text{Ru}$), this demonstrates that sample distillation does not cause non-exponential mass fractionation effects.

Isobaric interferences

The accuracy of the Ru isotopic measurements could be compromised by isobaric interferences from Mo, Pd, Zr and Ni argide species, or potential effects relating to remaining S in the analysed sample

solutions. While interferences from Mo and Pd are simultaneously monitored and corrected for during the measurements⁴⁸, isobars from Zr and Ni argides are not. Owing to the design of the collector block and limited availability of collectors, Zr and Ni could not be simultaneously monitored during the measurements. However, Zr is very effectively separated from Ru by cation exchange chemistry. Hence, all analysed sample solutions (except for one digestion of sample 10-27) had Zr intensities indistinguishable from the background of the 0.28 M HNO₃ and the Ru standard solution. Only one analysed sample (10-27) had a slightly elevated Zr/Ru ratio of 0.0008 and, hence, its $\epsilon^{96}\text{Ru}$ value is slightly elevated due to an isobaric interference from ^{96}Zr that could not be corrected. The Zr contained in the one analysed sample solution most probably reflects a random contamination from the laboratory equipment used during sample preparation that was not observed for other samples.

In the case of Ni we noticed during the initial stage of the project that a few processed reference samples still contained considerable amounts of Ni after the cation chemistry. For these samples, even after further purification of Ru by distillation, smaller amounts of Ni (between 1 and 10 ng ml⁻¹) were observed in the sample solutions to be analysed. During the isotopic measurements Ni readily forms argide species in the plasma that interfere with Ru isobars⁵⁰. To assess potential effects from Ni argide species on the measured Ru isotope data, a 100 ng ml⁻¹ Ru standard solution was doped with varying amounts of Ni to yield concentrations between 0.2 pg ml⁻¹ and 50 ng ml⁻¹. The results of this test show that the measured Ru isotope compositions for 100 ng ml⁻¹ Ru solutions are not affected for samples with Ni/Ru ratios <0.01 (Extended Data Fig. 3a, b). For sample solutions with higher amounts of Ni, positive anomalies are observed, which are most pronounced for ^{98}Ru and to a lesser extent for ^{100}Ru . Other Ru isotope masses (^{102}Ru and ^{104}Ru), owing to the lower abundance of the higher mass Ni isotopes, are not significantly affected by Ni argide species. To avoid any interferences from Ni argides during the isotopic measurements, the final dilutions of all samples analysed in this study were carefully checked for their Ni contents before the analysis. The intensity of Ni, monitored by scanning the mass of ^{58}Ni , in the finally diluted sample solutions was indistinguishable from the background intensity observed for the Ru solution standard and for 0.28 M HNO₃ (10–30 mV on ^{58}Ni). These negligible amounts of Ni are insignificant and have no effect on the measured data. The minimal Ni background originates from the Ni cones of the experimental set-up.

To eliminate any potential effects of S in the analysed solutions on the isotopic measurements, the S from the crushed NiS beads was almost completely removed by evaporation as H₂S gas during dissolution of the beads with concentrated HCl. After further purification of Ru by column chemistry and distillation, the S contents in the final sample solutions were <25 ng ml⁻¹ for all analysed samples. Tests with S-doped Alfa Aesar Ru standard solutions showed that even if a 100 ng ml⁻¹ Ru standard solution contains large excesses of S (S/Ru = 5), the accuracy of the Ru isotope measurements is not compromised (Extended Data Fig. 3c, d).

Nuclear field shift

Previous studies have shown that mass-independent Ru isotope anomalies could be caused by nuclear field shift-induced fractionation effects⁵¹. In meteorites and their components, such effects could be a primary feature resulting from evaporation/condensation processes. However, experimental studies have shown that mass-independent effects could also be generated in the laboratory during sample preparation⁵¹. In this section, we explore the potential effects of nuclear field shift-induced fractionation of Ru isotopes. These fractionations can be predicted on the basis of differences in the mean-squared nuclear charge radii between nuclides of a given element. The resulting effects on the measured Ru isotopic composition can be calculated in ϵ units using the following equation⁵¹:

$$\epsilon_{m_i} = \left(\delta r_{m_i, m_i}^2 - \frac{m_2(m_i - m_1)}{m_i(m_2 - m_1)} \delta r_{m_i, m_2}^2 \right) \alpha$$

where m_1 and m_2 are the atomic masses of the two isotopes of an element chosen for internal normalization, m_i refers to the atomic mass of another isotope indexed with variable i , δr^2 denotes the difference in the mean-squared nuclear charge radii of the respective isotope pair and α is an adjustable parameter that determines the magnitude of mass-independent fractionation, which is a function of temperature T as $1/T$. In plots of $\epsilon^{102}\text{Ru}$ and $\epsilon^{96}\text{Ru}$ versus $\epsilon^{100}\text{Ru}$ (Extended Data Fig. 4), the slopes calculated for the nuclear field shift fractionation are clearly distinct from the slope predicted by a variation in s-process Ru nuclides. The Ru isotopic composition obtained for Eoarchean southwest Greenland rocks does not plot on the calculated slope for nuclear field shift but instead plots on the s-process mixing line. As such, the anomalies identified in the southwest Greenland rocks cannot be explained by nuclear field shift-induced fractionation and therefore reflect isotope anomalies of nucleosynthetic origin.

Fissiogenic Ru

The spontaneous fission of uranium has been shown to produce Ru nuclides with relative abundances that are distinct from naturally occurring Ru (ref. ⁵²). Fissiogenic Ru primarily consists of ^{99}Ru (33.4%), ^{101}Ru (28.9%), ^{102}Ru (24.7%) and ^{104}Ru (12.4%)⁵³. The presence of an inherited fission-produced fraction of Ru in a rock sample would induce a characteristic isotope anomaly pattern that would be distinct from anomalies of nucleosynthetic or nuclear field shift origin. Because ^{96}Ru and ^{100}Ru are not a significant fission product⁵³, the presence of an inherited fraction of fissiogenic Ru in a rock sample would cause negative $\epsilon^{96}\text{Ru}$ and $\epsilon^{100}\text{Ru}$ anomalies. These are not observed in any of the analysed samples. On the other hand, a deficit of such an inherited fissiogenic Ru component would yield positive $\epsilon^{96}\text{Ru}$ and $\epsilon^{100}\text{Ru}$ anomalies, which are also not observed. This is shown in Extended Data Fig. 4b, where samples with an excess or a deficit of such a fissiogenic Ru components would fall on a mixing line with a distinct slope. Hence, the isotopic composition of the samples with a positive $\epsilon^{100}\text{Ru}$ anomaly cannot be explained by either an excess or a deficit of fissiogenic Ru nuclides.

Data availability

The data that support the findings of this study are available from the EarthChem library (<https://doi.org/10.1594/IEDA/111462>). Source data for Figs. 1–3 and Extended Data Fig. 1–4 are provided with the paper.

35. Nutman, A. P., McGregor, V. R., Friend, C. R. L., Bennett, V. C. & Kinny, P. D. The Itsaq Gneiss Complex of southern West Greenland; the world's most extensive record of early crustal evolution (3900–3600 Ma). *Precamb. Res.* **78**, 1–39 (1996).
36. Friend, C., Bennett, V. & Nutman, A. Abyssal peridotites > 3,800 Ma from southern West Greenland: field relationships, petrography, geochronology, whole-rock and mineral chemistry of dunite and harzburgite inclusions in the Itsaq Gneiss Complex. *Contrib. Mineral. Petrol.* **143**, 71–92 (2002).
37. Nutman, A. P., Friend, C. R. L., Horie, K. & Hidaka, H. in *Developments in Precambrian Geology* Vol. 15 (eds van Kranendonk, M. J. et al.) 187–218 (Elsevier, 2007).
38. Rollinson, H., Appel, P. W. U. & Frei, R. A. Metamorphosed, Early Archaean chromitite from West Greenland: implications for the genesis of Archaean anorthositic chromitites. *J. Petrol.* **43**, 2143–2170 (2002).
39. Coggon, J. A., Luguet, A., Nowell, G. M. & Appel, P. W. U. Hadean mantle melting recorded by southwest Greenland chromitite ^{186}Os signatures. *Nat. Geosci.* **6**, 871–874 (2013).
40. Friend, C. R. L. & Nutman, A. P. Eoarchean ophiolites? New evidence for the debate on the Isua supracrustal belt, southern West Greenland. *Am. J. Sci.* **310**, 826–861 (2010).
41. Friend, C. R. L. & Nutman, A. P. Dunites from Isua, Greenland: a ca. 3720 Ma window into subcrustal metasomatism of depleted mantle. *Geology* **39**, 663–666 (2011).
42. Szilas, K. et al. Highly refractory Archaean peridotite cumulates: petrology and geochemistry of the Seqi Ultramafic Complex, SW Greenland. *Geosci. Front.* **9**, 689–714 (2018).
43. Van Kranendonk, M. J., Hugh Smithies, R., Hickman, A. H., Wingate, M. T. D. & Bodorkos, S. Evidence for Mesoarchean (~3.2 Ga) rifting of the Pilbara Craton: the missing link in an early Precambrian Wilson cycle. *Precamb. Res.* **177**, 145–161 (2010).

44. Savard, D., Barnes, S.-J. & Meisel, T. Comparison between nickel-sulfur fire assay Te co-precipitation and isotope dilution with high-pressure asher acid digestion for the determination of platinum-group elements, rhenium and gold. *Geostand. Geoanal. Res.* **34**, 281–291 (2010).
45. Kaufmann, F. E. D. et al. Variations in composition, texture, and platinum group element mineralization in the Lower Group and Middle Group chromitites of the Northwestern Bushveld Complex, South Africa. *Econ. Geol.* **114**, 569–590 (2019).
46. Rizo, H. et al. Early Earth differentiation investigated through ^{142}Nd , ^{182}W , and highly siderophile element abundances in samples from Isua, Greenland. *Geochim. Cosmochim. Acta* **175**, 319–336 (2016).
47. Rehkämper, M. & Halliday, A. N. Development and application of new ion-dash-exchange techniques for the separation of the platinum group and other siderophile elements from geological samples. *Talanta* **44**, 663–672 (1997).
48. Fischer-Gödde, M., Burkhardt, C., Kruijer, T. S. & Kleine, T. Ru isotope heterogeneity in the solar protoplanetary disk. *Geochim. Cosmochim. Acta* **168**, 151–171 (2015).
49. Maréchal, C. N., Télouk, P. & Albarède, F. Precise analysis of copper and zinc isotopic compositions by plasma-source mass spectrometry. *Chem. Geol.* **156**, 251–273 (1999).
50. Becker, J. S., Seifert, G., Saprykin, A. I. & Dietze, H.-J. Mass spectrometric and theoretical investigations into the formation of argon molecular ions in plasma mass spectrometry. *J. Anal. At. Spectrom.* **11**, 643–648 (1996).
51. Fujii, T., Moynier, F., Telouk, P. & Albarède, F. Mass-Independent isotope fractionation of molybdenum and ruthenium and the origin of isotopic anomalies in Murchison. *Astrophys. J.* **647**, 1506–1516 (2006).
52. Hidaka, H. & Masuda, A. Isotopic search for spontaneous fission-produced ruthenium, silver and tellurium in uraninite. *Chem. Geol.* **106**, 187–195 (1993).
53. Hidaka, H. & Holliger, P. Geochemical and neutronic characteristics of the natural fossil fission reactors at Oklo and Bangombé, Gabon. *Geochim. Cosmochim. Acta* **62**, 89–108 (1998).
54. Chen, K. et al. Platinum-group element abundances and Re–Os isotopic systematics of the upper continental crust through time: evidence from glacial diamictites. *Geochim. Cosmochim. Acta* **191**, 1–16 (2016).
55. Schoenberg, R., Kruger, F. J., Nägler, T. F., Meisel, T. & Kramers, J. D. PGE enrichment in chromitite layers and the Merensky Reef of the western Bushveld Complex; a Re–Os and Rb–Sr isotope study. *Earth Planet. Sci. Lett.* **172**, 49–64 (1999).
56. Day, J. M. D., Brandon, A. D. & Walker, R. J. Highly siderophile elements in Earth, Mars, the Moon, and asteroids. *Rev. Mineral. Geochem.* **81**, 161–238 (2016).

Acknowledgements We thank F. Wombacher and A. Katzemich for support in the laboratory and K. Tani and I. Nishio for help during fieldwork. This research was supported through DFG grant number FI 1704/5-1 within the priority programme SPP 1833 ‘Building a Habitable Earth’ to M.F.-G., by the European Commission through ERC grant number 669666 ‘Infant Earth’ to C.M., by the Carlsberg Foundation grant number CF18-0090 to K.S., and by Kanazawa University “SAKIGAKE 2018” to T.M. H.S. publishes with the permission of the Executive Director, Geological Survey of Western Australia.

Author contributions M.F.-G. and C.M. designed the project. M.F.-G., B.-M. E. and N.M. developed the analytical method and obtained the Ru isotope data. M.F.-G. wrote the manuscript. All authors contributed to the discussion of the results and editing of the manuscript.

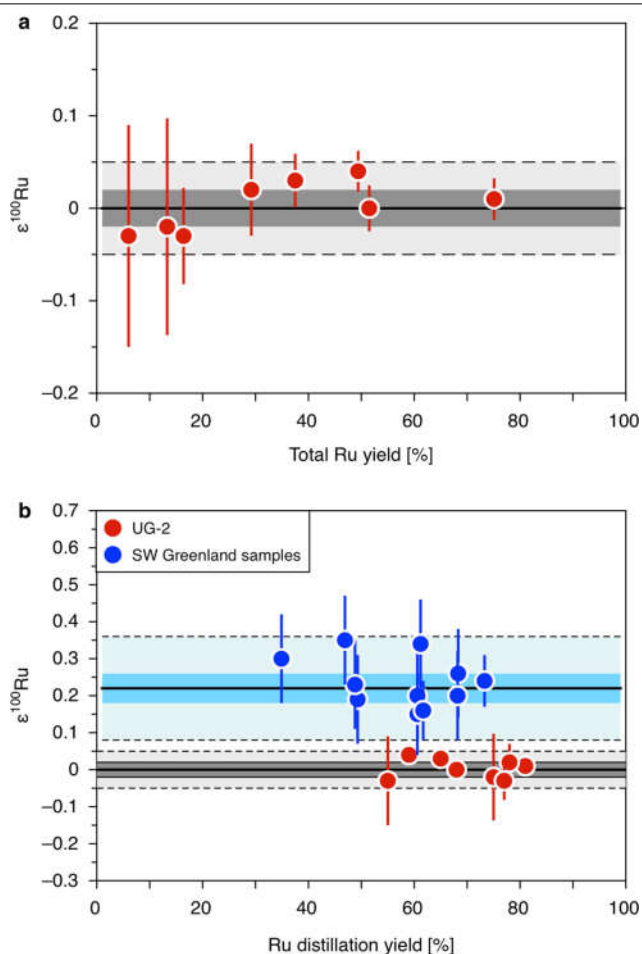
Competing interests The authors declare no competing interests.

Additional information

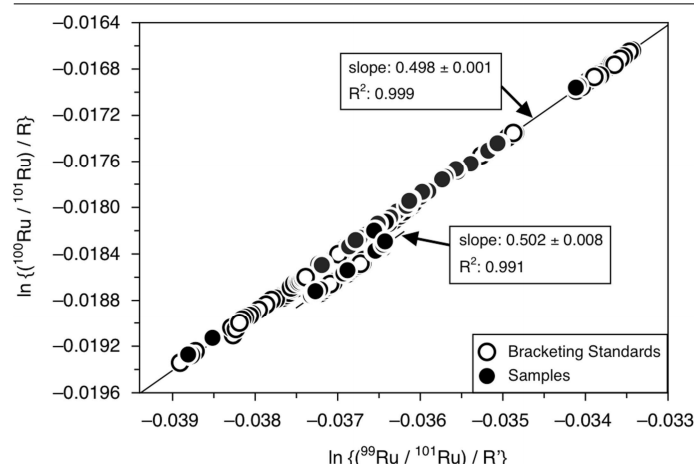
Correspondence and requests for materials should be addressed to M.F.-G.

Peer review information *Nature* thanks Katherine Bermingham and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

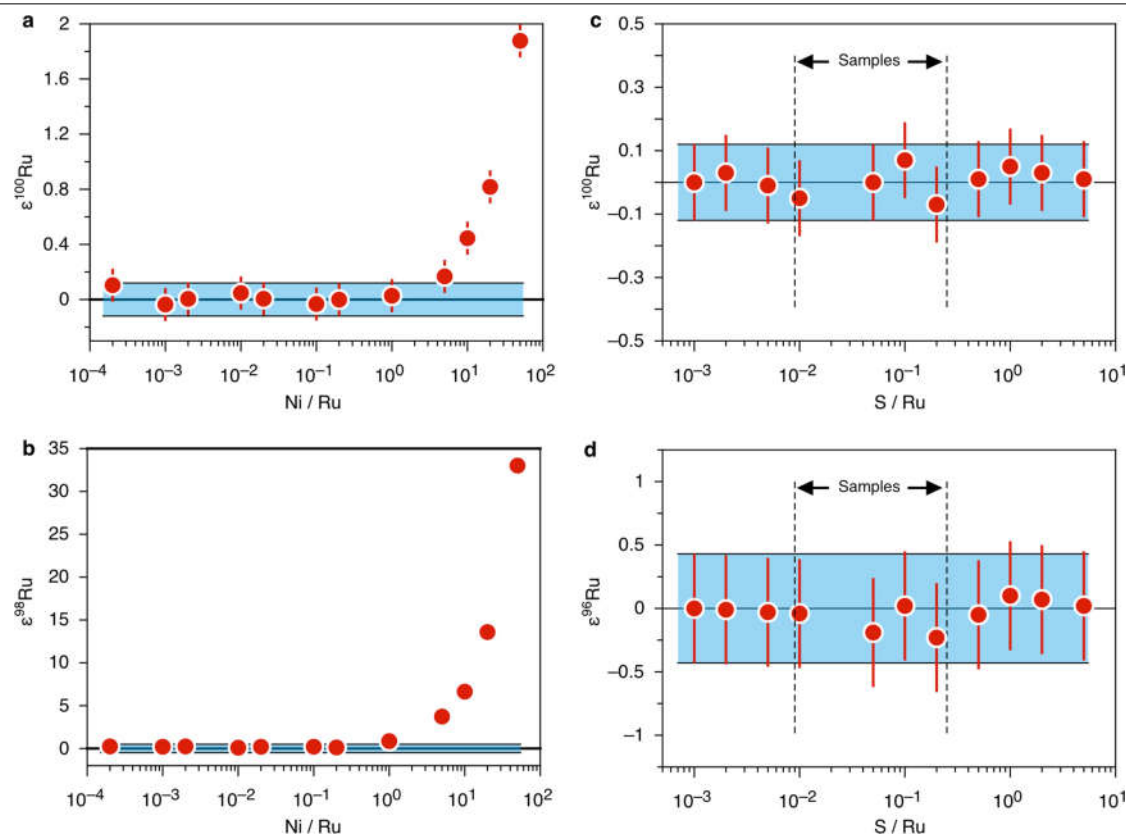
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Ruthenium yields of the total analytical procedure and Ru yields from the distillation plotted against $\epsilon^{100}\text{Ru}$ data for replicate digestions of UG-2 and southwest Greenland samples. **a, b, The accuracy of the $\epsilon^{100}\text{Ru}$ data are not affected by the Ru yield of the total analytical procedure (**a**) or the Ru yield of the distillation (**b**). The grey and blue areas represent the 95% confidence intervals, and the light grey and light blue areas limited by dashed lines indicate the 2 s.d. uncertainty as stated for the calculated mean values of the Bushveld igneous complex and the IGC (Extended Data Table 1). Note that it was not possible to determine the Ru distillation yield for all replicates of the analysed samples in **b** for the southwest Greenland samples.**



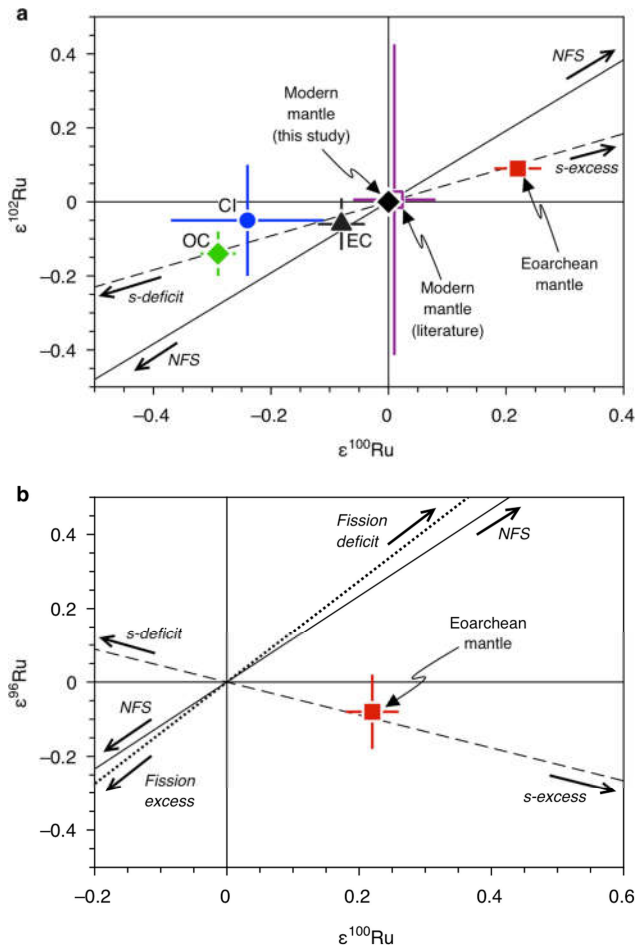
Extended Data Fig. 2 | Ruthenium isotope plot showing the \ln values of the measured raw ratios for $^{99}\text{Ru}/^{101}\text{Ru}$ and $^{100}\text{Ru}/^{101}\text{Ru}$ obtained for 100 ng ml^{-1} Ru sample solutions and associated Alfa Aesar Ru bracketing standards. The measured isotope ratios are shown as raw ratios uncorrected for mass-dependent fractionation and are normalized to reference ratios R and R' . Two distinct mass fractionation lines can be observed. The slopes for both lines are indistinguishable within error and are in very good agreement with the predicted slope of 0.5, which would be expected if the exponential law could accurately correct the mass-dependent fractionation (see Methods for details). Most importantly, the samples purified by distillation fall on the same respective mass fractionation line as their associated Alfa Aesar bracketing standards. This clearly demonstrates that the Ru distillation does not induce any unaccounted mass-fractionation effects for the samples in comparison with the bracketing standards. The shift observed for samples and associated standards plotting on a distinct mass fractionation array was caused by a maintenance on the mass spectrometer in May 2019, during which a Faraday cup was changed. However, this does not affect the accuracy of the isotopic data because the analysed samples and their associated bracketing standards are shifted by the same magnitude, and the isotopic data are expressed as ϵ values from the bracketing standards.



Extended Data Fig. 3 | Ruthenium isotope data obtained for 100 ng ml⁻¹ Alfa Aesar Ru standard solutions doped with variable amounts of Ni and S.

a, b, The effect of Ni argide interferences on the measured $\epsilon^{100}\text{Ru}$ (**a**) and $\epsilon^{98}\text{Ru}$ (**b**) isotopic compositions of a 100 ng ml⁻¹ Ru standard solution doped with varying amounts of Ni. The accuracy of the measured isotopic compositions is not affected for samples with $\text{Ni} / \text{Ru} < 10^{-2}$. **c, d,** High amounts of S ($\text{S} / \text{Ru} = 5$) in

the analysed sample solutions have no effect on the accuracy of the measured $\epsilon^{100}\text{Ru}$ (**c**) and $\epsilon^{98}\text{Ru}$ (**d**) data. Vertical dashed lines in **c** and **d** indicate the range of S / Ru ratios in the analysed samples. The blue areas indicate the external reproducibility of the method as defined by 2 s.d. of the replicate digestions and repeated analysis of the UG-2 reference sample (see Methods).



Extended Data Fig. 4 | Ruthenium isotope plots illustrating systematic compositional differences between EC, OC and CI, modern mantle and the pre-late-veer mantle in comparison with mixing lines calculated between the modern mantle composition and isotopic variations caused by variable contributions of s-process Ru nuclides, a fissionogenic Ru component and variations of nuclear field-shift induced isotope fractionations.

a, Ruthenium isotope data for the Eoarchean mantle of southwest Greenland, the modern mantle and the three types of carbonaceous chondrite in comparison with a mixing line calculated between the composition of the modern mantle ($\epsilon^{102}\text{Ru} = \epsilon^{100}\text{Ru} = 0$) and an s-process component³⁴ (dashed line) and the slope calculated for a nuclear field shift-induced (NSF) fractionation⁵¹ (solid black line). **b**, Ruthenium isotope composition of the Eoarchean mantle in comparison with mixing lines calculated between the composition of the modern mantle ($\epsilon^{96}\text{Ru} = \epsilon^{100}\text{Ru} = 0$) and a fissionogenic Ru component⁵³ (dotted line), an s-process component³⁴ (dashed line) and the slope calculated for an NSF isotope fractionation⁵¹ (solid black line).

Extended Data Table 1 | Ruthenium isotope data for ultramafic rocks

Sample*	Rock†	n‡	ε ⁹⁶ Ru§	±	ε ⁹⁶ Ru§	±	ε ¹⁰⁰ Ru§	±	ε ¹⁰² Ru§	±	ε ¹⁰⁴ Ru§	±
Greenland												
Itsaq gneiss complex (IGC)												
<i>Narssaq ultramafic body (NUB), 3.8 Gyr</i>												
10_9	H	1	0.13	0.43	-0.04	0.49	0.19	0.12	0.10	0.16	0.17	0.30
replicate	H	2	0.28	0.43	0.42	0.49	0.30	0.12	0.13	0.16	-0.01	0.30
10_11	D	2	-0.07	0.43	0.29	0.49	0.34	0.12	0.08	0.16	-0.10	0.30
<i>South of the Isua supracrustal belt (SOISB), 3.8 Gyr</i>												
10_27	H	1	(1.19) [¶]	0.43	0.33	0.49	0.23	0.12	0.03	0.16	0.04	0.30
replicate		1	0.03	0.43	0.58	0.49	0.23	0.12	0.18	0.16	0.00	0.30
<i>Ujaragssuit Nunât complex, 3.8 Gyr</i>												
194856	Chr	6	-0.04	0.20	0.18	0.17	0.17	0.09	0.06	0.06	-0.10	0.08
replicate		5	-0.27	0.29	0.04	0.35	0.15	0.11	0.07	0.04	-0.03	0.13
replicate		3	-0.25	0.43	0.40	0.49	0.35	0.12	0.13	0.16	-0.11	0.30
194857	Chr	10	-0.20	0.09	0.06	0.18	0.19	0.03	0.12	0.06	0.00	0.08
replicate		5	-0.17	0.22	0.17	0.31	0.16	0.08	0.06	0.17	-0.08	0.16
replicate		11	-0.09	0.20	0.10	0.18	0.13	0.04	0.05	0.05	-0.11	0.11
194882B	Chr	3	0.03	0.43	0.30	0.49	0.19	0.12	0.07	0.16	0.13	0.30
194884C	Chr	7	-0.17	0.11	0.06	0.14	0.24	0.07	0.09	0.09	-0.08	0.09
<i>Isua supracrustal belt (ISB), 3.7 Gyr</i>												
194907	D	1	-0.23	0.43	0.24	0.49	0.26	0.12	0.10	0.16	-0.20	0.30
Mean IGC (n=14)			-0.08	0.10	0.22	0.10	0.22	0.04	0.09	0.02	-0.03	0.06
2 s.d. IGC				0.33		0.35		0.14		0.08		0.20
Seqi ultramafic complex, >3.0 Gyr												
186466	Chr	2	-0.44	0.43	-0.10	0.49	0.20	0.12	0.04	0.16	-0.29	0.30
replicate		2	-0.14	0.43	0.60	0.49	0.13	0.12	0.04	0.16	-0.13	0.30
186479	Chr	3	0.10	0.43	0.30	0.49	0.20	0.12	0.08	0.16	-0.17	0.30
Australia												
Pilbara craton, Warawoona Group, 3.5-3.2 Gyr												
Pil 16-61	Chr	3	-0.03	0.43	-0.11	0.49	0.05	0.12	0.01	0.16	-0.02	0.30
replicate		3	-0.25	0.43	-0.17	0.49	0.05	0.12	-0.06	0.16	-0.05	0.30
Canada												
Abitibi greenstone belt, 2.7 Gyr												
OKUM	Kom	1	0.24	0.43	0.58	0.49	0.07	0.12	-0.03	0.16	0.08	0.30
South Africa												
Bushveld igneous complex, 2.05 Gyr												
UG-2	Chr	20	-0.12	0.08	-0.06	0.10	0.00	0.02	0.03	0.04	0.05	0.07
replicate		29	-0.12	0.08	0.03	0.08	0.01	0.02	-0.01	0.03	-0.06	0.06
replicate		19	-0.11	0.10	0.13	0.11	0.04	0.02	0.04	0.04	0.05	0.07
replicate		11	-0.11	0.15	0.18	0.11	0.02	0.05	0.01	0.04	-0.02	0.10
replicate		5	-0.08	0.47	-0.19	0.44	-0.02	0.12	-0.02	0.11	0.06	0.23
replicate		2	-0.15	0.43	-0.34	0.49	-0.03	0.12	-0.03	0.16	0.08	0.30
replicate		6	-0.31	0.25	0.13	0.31	-0.03	0.05	-0.02	0.07	0.08	0.10
replicate		11	0.02	0.12	-0.01	0.16	0.03	0.03	-0.02	0.05	-0.04	0.09
LG-6	Chr	4	-0.20	0.28	0.05	0.32	0.01	0.09	-0.01	0.09	-0.04	0.18
Mean Bushveld (n=9)			-0.13	0.07	-0.01	0.13	0.00	0.02	0.00	0.02	0.02	0.04
2 s.d. Bushveld				0.18		0.33		0.05		0.05		0.12

*The term 'replicate' indicates that the Ru isotope data were obtained for a replicate digestion of the same sample powder.

†H, harzburgite; D, dunite; Chr, chromitite; Kom, komatiite.

‡Number of analyses of the same sample solution.

§Ruthenium isotope data are internally normalized to ⁹⁹Ru/¹⁰¹Ru using the exponential law and are reported as ε unit (0.01%) deviations from the terrestrial bracketing standard:

εⁱRu = ((ⁱRu/¹⁰¹Ru)_{sample} / (ⁱRu/¹⁰¹Ru)_{standard} - 1) × 10⁴, where i = 96, 98, 100, 102 or 104. For samples measured n < 4 times, the quoted errors reflect the external uncertainty as defined by the standard deviation (2 s.d.) of replicate digestions and repeated analysis of a reference sample (see Methods). Uncertainties for samples measured n ≥ 4 times are given as 95% confidence intervals of the mean calculated as (s.d. × t_{0.95,n-1})/√n.

^{||}Calculated averages for samples from IGC and Bushveld Complex with 95% confidence interval uncertainties calculated as (s.d. × t_{0.95,n-1})/√n and 2 s.d. uncertainties.

[¶]Elevated ε⁹⁶Ru value due to an isobaric Zr interference that could not be corrected during the isotopic measurement (see Methods). Consequently, this value was not included for the calculation of the IGC mean value.

Extended Data Table 2 | Ruthenium concentration data, details about the NiS procedure and Re–Os isotope data for ultramafic rocks

Sample	Rock*	Ru (ng g ⁻¹) [†]	NiS n x (g) [‡]	¹⁸⁷ Os/ ¹⁸⁸ Os [§]	¹⁸⁷ Re/ ¹⁸⁸ Os [§]	Age (Gyr)	¹⁸⁷ Os/ ¹⁸⁸ Os _{ini.} [¶]	yOs [#]	Ref. ^{**}
Greenland									
Itsaq gneiss complex (IGC)									
<i>Narssaq ultramafic body (NUB), 3.8 Gyr</i>									
10_9	H	7.76	5 (8) x10g	0.1013	0.044	3.80	0.09845	-2.3	14
10_11	D	8.27	8 x10g	0.1013	0.042	3.80	0.09857	-2.2	14
<i>South of the Isua supracrustal belt (SOISB), 3.8 Gyr</i>									
10_27	H	3.13 - 6.80	5 (6) x10g	0.1010 - 0.1153	0.023 - 0.233	3.81	0.09942 - 0.10002	-5.3 to +5.8	14
<i>Ujaragssuit Nunât complex, 3.8 Gyr</i>									
194856	Chr	215	1 x5g	0.10471 - 0.10493	0.0008 - 0.0013	3.81	0.1046 - 0.1047	+3.9 to +4.0	33
194857	Chr	189	1 x5g	0.10490 - 0.10809	0.01424 - 0.02906	3.90	0.10395 - 0.10614	+3.9 to +6.1	38
194882B	Chr	12	6 x10g	0.104842 - 0.105590					39
194884C	Chr	23	4 x10g						
<i>Isua supracrustal belt (ISB), 3.7 Gyr</i>									
194907	D	2.43 - 7.06	7 x10g	0.10899 - 0.37064	0.049 - 2.96	3.80	0.1045 - 0.1772	+3.8 to +76.2	23
		2.9 - 5.3		0.10245 - 0.13411	0.042 - 0.707	3.70	0.089 - 0.101	-12.8 to +0.2	46
Seqi ultramafic complex, >3.0 Gyr									
186466	Chr	53.83	2 x5g						42
186479	Chr	53.54	2 x5g						42
Australia									
Pilbara craton, Warawoona Group, 3.5-3.2 Gyr									
Pil 16-61	Chr	115	1 x5g	0.10517 - 0.11926	0.0281 - 0.1688	3.46	0.0979 - 0.1130	-5.2 to +9.6	33
Canada									
Abitibi greenstone belt, 2.7 Gyr									
OKUM	Kom	4.25	4 x10g	0.269					44, 54
South Africa									
Bushveld igneous complex, 2.05 Gyr									
UG-2	Chr	760	1 x5g	0.1483 - 0.1489	0.0916				45, 55
LG-6	Chr	270 - 430	1 x5g	0.1254 - 0.1265	0.101 - 0.117				45, 55

*H, harzburgite; D, dunite; Chr, chromitite; Kom, komatiite.

[†]Ruthenium concentrations reported for the same samples or samples from the same locality. Italicized values indicate Ru concentrations determined in this study as described in the Methods.

[‡]Number of individual NiS digestions (n) and amount of sample material (in g) taken for each single NiS digestion processed for each sample and replicate sample analysis. In cases where different numbers of NiS digestions were processed for a replicate sample analysis, this is indicated by the number in parentheses. If not stated otherwise, the same number of NiS digestions was used for all replicates.

[§]Re–Os isotope data reported for the same samples or for samples of similar composition from the same locality. Note that the different ranges reported for chromitite samples from the Ujaragssuit Nunât complex were not obtained for the same samples and represent data reported by different previous studies for chromitites from the same locality or sampling area.

^{||}Age used for calculation of the initial ¹⁸⁷Os/¹⁸⁸Os isotope composition as quoted in literature.

[¶]Initial ¹⁸⁷Os/¹⁸⁸Os isotope composition calculated at the time of emplacement of the samples.

[#]yOs values refer to the relative deviation (in %) of the calculated initial ¹⁸⁷Os/¹⁸⁸Os isotope compositions of samples relative to the chondritic Os isotope composition at the time of emplacement of the samples.

**Data from refs. ^{14,23,33,38,39,42,44–46,54,55} as indicated.

Extended Data Table 3 | Parameters for mixing model shown in Fig. 3 and resulting $\epsilon^{100}\text{Ru}$ values for pre-late-veener mantle endmember composition

	$\epsilon^{100}\text{Ru}$	\pm	Ru (ng g ⁻¹)	$\epsilon^{100}\text{Ru}$ in pre-LV mantle	$\epsilon^{100}\text{Ru}$ in pre-LV mantle min	$\epsilon^{100}\text{Ru}$ in pre-LV mantle max
Chondrite groups						
Enstatite	-0.08	0.04	818 ⁵⁶	0.31	0.16	0.47
Ordinary	-0.29	0.03	882 ⁵⁶	1.14	1.02	1.26
Carbonaceous average	-0.90	0.61	838 ⁵⁶	3.56	1.19	4.74
Carbonaceous average	-0.90	0.12	838 ⁵⁶	3.56	3.08	4.03
CI	-0.24	0.13	637.4 ⁵⁶	1.00	0.46	1.55
CM	-0.69	0.38	817 ⁵⁶	2.90	1.30	4.49
Terrestrial mantle						
Modern mantle*	0.00	0.02	7.0 ¹⁶			
Eoarchean SW Greenland mantle†	0.22	0.04				
Pre-late veneer mantle			1.4			

*The composition of the modern mantle corresponds to the mean value calculated for samples from the Bushveld complex (Extended Data Table 1).

†The Eoarchean southwest Greenland mantle composition represents the mean value calculated for samples from the IGC (Extended Data Table 1). Data from refs. ^{16,56}.

Hummingbird-sized dinosaur from the Cretaceous period of Myanmar

<https://doi.org/10.1038/s41586-020-2068-4>

Received: 12 September 2019

Accepted: 22 January 2020

Published online: 11 March 2020

 Check for updates

Lida Xing^{1,2,11}, Jingmai K. O'Connor^{3,4,11}✉, Lars Schmitz^{5,6,11}, Luis M. Chiappe⁵, Ryan C. McKellar^{7,8}, Qiru Yi⁹ & Gang Li^{9,10,11}

Skeletal inclusions in approximately 99-million-year-old amber from northern Myanmar provide unprecedented insights into the soft tissue and skeletal anatomy of minute fauna, which are not typically preserved in other depositional environments^{1–3}. Among a diversity of vertebrates, seven specimens that preserve the skeletal remains of enantiornithine birds have previously been described^{1,4–8}, all of which (including at least one seemingly mature specimen) are smaller than specimens recovered from lithic materials. Here we describe an exceptionally well-preserved and diminutive bird-like skull that documents a new species, which we name *Oculudentavis khaungraae* gen. et sp. nov. The find appears to represent the smallest known dinosaur of the Mesozoic era, rivalling the bee hummingbird (*Mellisuga helenae*)—the smallest living bird—in size. The *O. khaungraae* specimen preserves features that hint at miniaturization constraints, including a unique pattern of cranial fusion and an autapomorphic ocular morphology⁹ that resembles the eyes of lizards. The conically arranged scleral ossicles define a small pupil, indicative of diurnal activity. Miniaturization most commonly arises in isolated environments, and the diminutive size of *Oculudentavis* is therefore consistent with previous suggestions that this amber formed on an island within the Trans-Tethyan arc¹⁰. The size and morphology of this species suggest a previously unknown bauplan, and a previously undetected ecology. This discovery highlights the potential of amber deposits to reveal the lowest limits of vertebrate body size.

Aves, Linnaeus 1758

Oculudentavis khaungraae gen. et sp. nov. (Fig. 1)

Holotype. Hupoge Amber Museum (HPG)-15-3, a complete skull preserved in amber; the block measures 31.5 mm × 19.5 mm × 8.5 mm and weighs 2.84 g.

Etymology. The generic name *Oculudentavis* is derived from the Latin *oculus* (eye), *dentes* (teeth) and *avis* (bird). The species name *khaungraae* is from Khaung Ra, who donated the specimen to the Hupoge Amber Museum.

Locality and horizon. Cenomanian age, 98.8 ± 0.6 million years ago¹¹. Angbamo site, Tanai township (Myitkyina district, Hukawng valley, Kachin province), northern Myanmar.

Diagnosis. Very small bird with the following autapomorphies: jugal process of maxilla reaches caudally to the level of mid-orbit; jugal bar cross-section strongly angled dorsolaterally–ventromedially; triangular, dorsolaterally oriented coronoid process on mandible; and 23 teeth in the upper jaw, 4 of which are located beneath the orbit. The taxon can be further diagnosed by the unique combination of the following

features: slender rostrum; antorbital fenestra reduced or absent; and scleral ossicles spoon-shaped, longer (from external to internal margin) than they are wide (distance between adjacent ossicles). Although the medial articulations between the premaxillae and dentaries are discernible, high-resolution synchrotron computed tomography scans reveal that the bones are partially fused. We consider HPG-15-3 to be skeletally mature or nearly so (Supplementary Information), and consider its small size and an unusual pattern of cranial fusion (with the premaxilla, maxilla and nasal fused into a single unit) to be diagnostic features of *O. khaungraae*.

Description

HPG-15-3 is mesorostrine (Fig. 1, Extended Data Fig. 1). The jugals are bowed laterally and the skull width is greatest at the caudal margin of the orbit. The upper jaw is wider than the lower jaw, the dentition of which occludes with the ventral surface of the secondary palate.

The cranial half of the rostrum is imperforate. Nutrient foramina are present on the lateral surface of the premaxillae, maxillae and dentaries. The premaxillae are partially fused medially, such that the

¹State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing, China. ²School of the Earth Sciences and Resources, China University of Geosciences, Beijing, China. ³Key Laboratory of Vertebrate Evolution and Human Origins of the Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Beijing, China. ⁴Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, China. ⁵Dinosaur Institute, Natural History Museum of Los Angeles County, Los Angeles, CA, USA. ⁶W. M. Keck Science Department, Claremont McKenna, Scripps and Pitzer Colleges, Claremont, CA, USA. ⁷Royal Saskatchewan Museum, Regina, Saskatchewan, Canada.

⁸Biology Department, University of Regina, Regina, Saskatchewan, Canada. ⁹Beijing Advanced Sciences and Innovation Center, Chinese Academy of Sciences, Beijing, China. ¹⁰Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China. ¹¹These authors contributed equally: Lida Xing, Jingmai K. O'Connor, Lars Schmitz, Gang Li. ✉e-mail: jingmai@ivpp.ac.cn

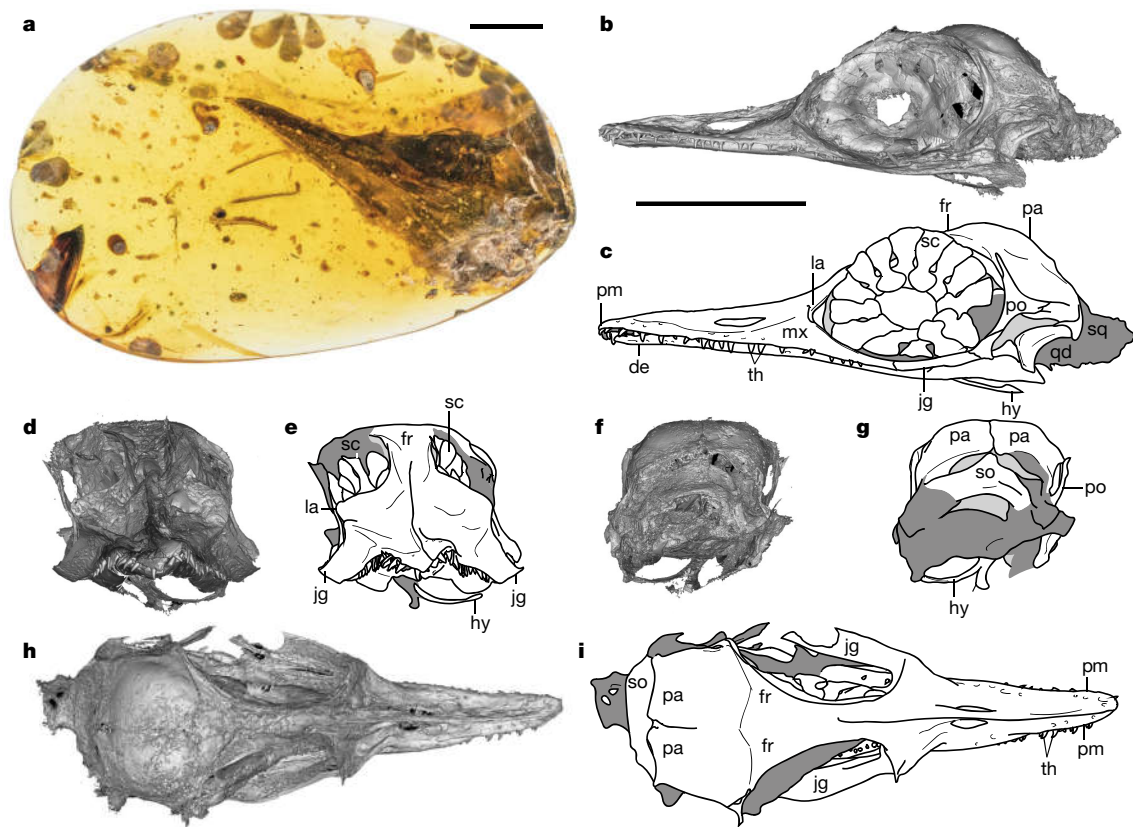


Fig. 1 | Photograph, computed tomography scans and interpretive drawings of the HPG-15-3 holotype of *O. khaungraae*. **a**, Photograph of the amber piece with skull ventrolaterally exposed. **b**, **c**, Scan (**b**) and drawing (**c**), left lateral view. **d**, **e**, Scan (**d**) and drawing (**e**), rostral view. **f**, **g**, Scan (**f**) and drawing (**g**), occipital

view. **h**, **i**, Scan (**h**) and drawing (**i**), dorsal view. de, dentary; fr, frontal; hy, hyoid bone (or bones); jg, jugal; la, lacrimal; mx, maxilla; pa, parietal; pm, premaxilla; po, postorbital; qd, quadrate; sc, scleral ossicle; so, supraoccipital; sq, squamosal; th, teeth. Scale bars, 5 mm; longer scale bar below **b** applies to **b**–**i**.

right premaxilla can be clearly distinguished from the left in rostral or dorsal view. Articulation with the maxillae cannot be discerned, which suggests that these elements were fused to one another. The frontal processes of the premaxillae are elongate, and apparently articulate deeply with the frontals. The external nares have rounded rostral margins and tapered caudal margins.

The maxilla is long and toothed along its entire length. Caudal to the nares, the maxilla expands dorsolaterally until it reaches the rostral margin of the orbit, forming a long and broad nasal process. The jugal ramus of the maxilla extends beyond the rostral margin of the orbit, articulating with the ventral surface of the jugal (Extended Data Fig. 2). Medially, the jugal ramus of the maxilla contacts the pterygoid—whether this is through an ectopterygoid or not is unclear—and ends just caudal to this contact. The lacrimal has well-developed dorsomedial and ventral processes, and a very small dorsolateral process. The lacrimal is roughly trapezoidal in dorsal aspect; the rostral margin is wedged between the nasal medially, and the maxilla laterally (Fig. 1c). There is no clear evidence of an antorbital fenestra. The jugal and maxilla have a long, dorsoventrally overlapping articulation, such that the jugal forms the entirety of the concave, ventral margin of the orbit¹². The jugal is strongly compressed, bowed and transversally inclined such that its cross-section is angled dorsolaterally–medioventrally. The caudal end of the jugal tapers dorsocaudally, articulating with the jugal process of the postorbital.

The large, Y-shaped postorbital forms a complete infratemporal bar and demarcates the infratemporal and supratemporal fenestrae (Fig. 1). The quadrate is angled dorsocaudally–rostroventrally. The otic process forms a single articular surface (Extended Data Fig. 2c, d). The triangular orbital ramus is short and broad, articulating with the pterygoid medially.

The orbits are large and entirely occupied by massive scleral rings; their internal margins are medially displaced (that is, the internal margin is medial to the external margin, opposite to the *in vivo* condition in neornithines). The ring consists of 14 scleral ossicles (the modal number in saurians). The ossicles are large, and their length is twice that of their greatest width. The internal half of each ossicle is rounded and the external half is strap-like, expanded in width externally. As preserved, the internal diameter of the scleral ring measures slightly less than the long axis of the ossicles themselves. An interorbital septum is not visible and the mesethmoid appears to be absent.

The frontals are long, slender and medially fused (Fig. 1). The parietals are unfused to the frontals, which is typical of nonornithurine birds (Extended Data Fig. 3d). Medially, the parietals are fused only along the rostral half of their contact. The parietals are large, longer than they are wide and domed. Their caudal margins form a delicate nuchal crest. The triradiate squamosal is unfused to the braincase. The occipital table is poorly preserved, owing to the activity of bivalves¹³ (Extended Data Fig. 4g). The broad supraoccipital forms the dorsal margin of the large caudally opened foramen magnum (Fig. 1f, g). Short paraoccipital processes project sideways at roughly the level of the dorsal margin of the foramen magnum. Only the dorsal surface of the brain is clear in the endocast, and reveals well-developed cerebral hemispheres that—in dorsal view—are as wide as they are long (Extended Data Fig. 4).

The parasphenoidal lamina is large and flat (Extended Data Fig. 2a). The palatal configuration resembles that of palaeognaths in having rostrally projecting basipterygoid processes. The pterygoids are cranially forked into a short maxillary process and a longer palatine process. The tapered parasphenoidal rostrum extends rostrally just past the midpoint of the orbit. The rostral half of the palate is imperforate,

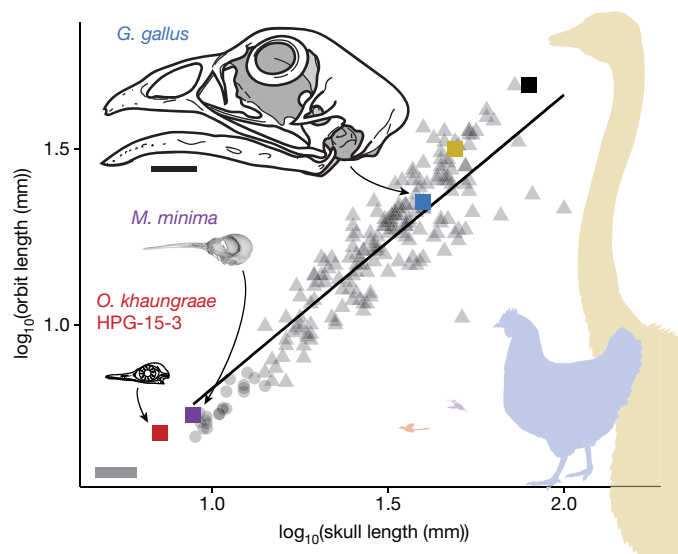


Fig. 2 | Proportions of the eye socket relative to the skull in HPG-15-3, compared to extant birds. The plot of \log_{10} -transformed orbit length (y-axis) versus \log_{10} -transformed skull length for extant bird species (grey triangles) ($n = 206$), with the phylogenetic generalized least squares line obtained for the majority-rule consensus tree for the Hackett backbone tree set (see ‘Scaling of eye socket and skull length’ sections in the Methods and Supplementary Information for details). The skull of HPG-15-3 (red square) is smaller than any of the extant birds that we included, but the eye socket is about as large as expected if the fitted line is extended towards HPG-15-3. The skull drawings of *M. minima* (purple square), *O. khaungraae* (red square) and *Gallus gallus* (blue square) are depicted to a relative scale (scale bar equals 10 mm). The yellow and black squares plot the ostrich (*Struthio camelus*) and the greater rhea (*Rhea americana*), respectively. The silhouettes depict the body size of *Oculudentavis* relative to *Mellisuga*, *Struthio* and *Gallus* (colours of the silhouettes correspond to their respective data points).

formed by the medially unfused palatal processes of the premaxillae (Extended Data Fig. 3c). The choanae appear to be level with the external nares. The preserved oral soft tissue bears a pattern of elongated papillae¹⁴ (Extended Data Fig. 2a, b).

The lower jaws are thin and nearly straight, with subparallel dorsal and ventral margins (Extended Data Fig. 5). The medial suture between dentaries is clearly visible; no predentary bone is present. The boundaries between postdentary bones cannot be discerned. A triangular, dorsolaterally oriented coronoid process is present at one quarter length from the caudal margin of the lower jaw. The bowed (that is, laterally concave) ceratobranchial elements are preserved with their cranial ends in contact with one another (Fig. 1e).

The upper jaw contains 23 teeth on the right side; 18 teeth are preserved on the left side. The rostralmost three teeth are somewhat procumbent. The apices of the rostralmost seven teeth are weakly curved lingually. The caudal teeth are conical and ventrally oriented. The mesial and distal margins of the teeth are gently carinated. The teeth lack basal constrictions and have expanded roots; they vary in size, crown height and spacing. The teeth are largest in the portion of the maxilla that is ventral to the external nares, and are smallest underneath the orbit. Each dentary had approximately 29 or 30 teeth, which resemble those in the upper jaw. The tooth geometry appears to be acrodont to pleurodont; no grooves or sockets are discernable.

Discussion

The skull of *O. khaungraae* reveals a general avian morphology, including a slender, tapering rostrum with retracted naris, an enlarged and well-defined eye socket, a short compressed postorbital region and a

dome-shaped skull roof. *O. khaungraae* appears to represent the smallest known Mesozoic dinosaur, rivalling the extant bee hummingbird—the smallest known dinosaur of all time—in size (Fig. 2). The length of the postrostral portion of the skull (measured from the contact between the nasal and frontal at the sagittal line to the caudal end of the cranium) is 7.1 mm, compared to 8.8 mm in the vervain hummingbird (*Mellisuga minima*)—which is only slightly larger than the bee hummingbird¹⁵. No other group of living birds features species with similarly small crania in adults (Fig. 2, Extended Data Fig. 6). The discovery of *Oculudentavis* highlights the presence of diminutive members of vertebrate faunas that possibly can only be studied through preservation in amber, which represents a taphonomic filter capturing the lowest end of the body-size spectrum. This discovery is also consistent with hypotheses that the Cretaceous Burmese amber from the Angbamo site formed in an island arc, as miniaturization most commonly arises in island environments^{10,16}.

HPG-15-3 displays morphologies that, to our knowledge, have not previously been observed in any bird; these morphologies include features that depart from the theropod condition altogether (for example, spoon-shaped scleral ossicles and acrodont to pleurodont dentition), as well as an unusual combination of traits that are primitive (for example, complete postorbital bar and extensive tooth row) and advanced (for example, an expanded imperforate portion of the rostrum and an orbit confluent with antorbital fenestra) for avians. Nearly all of these unusual morphologies can be interpreted as the effects of miniaturization, which is commonly associated with the reappearance of plesiomorphic morphologies, increased bone fusion and proportionally enlarged sensory organs (for example, the eye and middle ear)^{9,16,17}. Miniaturization is most commonly associated with paedomorphism that results from progenesis^{9,18}. However, the skull of *Oculudentavis* reveals no obvious paedomorphic features (for example, a proportionally large orbit and short rostrum), which suggests that miniaturization in this lineage may have been achieved by a reduction in growth rate⁹.

The degree of fusion between the skull bones that form each half of the rostrum and the pattern in which the sutures have closed is highly unusual among nonneornithine theropods. In early birds, the premaxillary bodies—followed by the dentaries—are consistently the first skull elements to fuse in both an ontogenetic and phylogenetic context, and these sutures are completely obliterated in mature specimens of some taxa (for example, Confuciusornithiformes and the enantiornithine *Gobipteryx*)^{19,20}. In Late Cretaceous ornithurines the mandibular bones remain unfused, although the premaxillae form a single element¹². In *Gobipteryx* and neornithines, the premaxillae fuse into a single element during embryonic development^{20,21}. The pattern of fusion in HPG-15-3 departs from this pattern and is distinct from all other known theropods, highlighting the enigmatic nature of *Oculudentavis*. This fusion pattern may be related to structural constraints that are imparted by miniaturization⁹ and to the predatory ecology inferred for this animal.

The scleral ring of *Oculudentavis* is very different from that preserved in any known Mesozoic dinosaur, in which the ossicles are typically nearly square-to-rectangular, narrow and demarcate a proportionally larger aperture (as seen, for example, in *Archaeopteryx*, *Sapeornis* and *Yixianornis*). In HPG-15-3, the scleral ring is very large and is formed by elongated spoon-shaped ossicles; a morphology similar to this is otherwise known only in lizards (for example, *Lacerta viridis*)²². The relatively small aperture defined by the scleral ring of *Oculudentavis* suggests unique visual capabilities compared to other Mesozoic theropods (Fig. 3), and indicates that the maximum size of the pupil was fairly small, limiting the amount of light entering the eye^{23,24}. Even accounting for a range of values, this morphology suggests a diurnal lifestyle in photopic light environments²⁵ (Fig. 3). Although both scleral rings appeared deformed, the elongation of the ossicles—together with the bowed morphology of the jugal bar—suggests that these elements would have defined a conical eye similar to that of some extant birds (Extended Data Figs. 7, 8) and lizards²². Nanoid taxa typically have enlarged eyes^{9,17} that result from negatively allometric scaling of the eyes to body size.

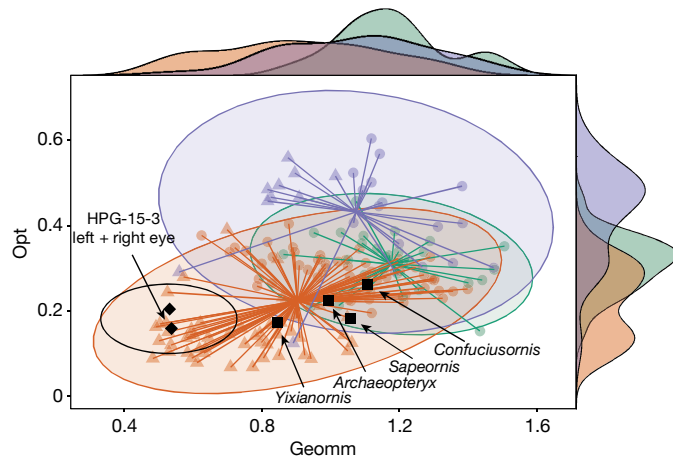


Fig. 3 | A star plot illustrating the functional morphospace of scleral-ring and orbit morphology of modern saurians in the context of their diel activity pattern. The x axis (labelled 'geommm') represents the \log_{10} transformed geometric mean of orbit length and the external and internal scleral-ring diameters, and the y axis (labelled 'opt') is formed by the optical ratio. Triangles indicate squamates ($n = 60$ species); circles indicate birds ($n = 104$ species). Ellipses represent 95% confidence areas for cathemeral (green), diurnal (orange) and nocturnal (purple) groups, drawn around the respective average x and y values for each group. Density plots illustrate the distribution of data along the x and y axes, delineated by groups of diel activity pattern. The retrodeformed eyes of HPG-15-3 are most similar in shape and size to the eyes of diurnal squamates. The black oval around the data points representing the left and right eyes of HPG-15-3 represent the 95% confidence ellipse, indicating the range of values estimated by our sensitivity analysis. The average of the geometric mean of all eye variables and the optical ratio for the retrodeformed left and right eyes were calculated and used to generate a distribution of possible combinations spanning $\pm 30\%$ of the left- and right-eye averages.

Oculudentavis follows this pattern and falls where expected in a linear regression of avian orbit size relative to skull length (Fig. 2, Extended Data Fig. 9). The laterally bulging eyes of *Oculudentavis*, as indicated by the angled morphology of the jugal bar, probably represent an alternative strategy for increasing eye size without further increasing the size of the orbit. Notably, *Oculudentavis* lacks the forward-facing position of the eyes of owls and living birds of prey; the eyes are laterally oriented such that binocular vision was absent, or largely limited. Given the unusual morphologies observed in HPG-15-3, extant analogues for the visual abilities of *Oculudentavis* may not exist.

Although miniaturization is often associated with edentulism⁹, *Oculudentavis* possesses an extensive dentition with unusual features. First, the acrodont to pleurodont geometry of tooth implantation differs from the thecodont condition in all other dinosaurs. Second, the tooth row extends ventral to the orbit; this condition is otherwise only observed among coelurosaur theropods in the Late Cretaceous ornithurine bird *Ichthyornis*^{12,26}. In most coelurosaurs (including Aves), the maxillary tooth row typically extends as far caudally as the rostral half of the antorbital fenestra (well cranial to the orbit²⁷), although the tooth row extends nearly to the jugal–lacrima contact in some troodontids²⁸ and in *Ichthyornis* one tooth is located caudal to the lacrimal and beneath the orbit. The tooth row is even more caudally extensive in *Oculudentavis*, which has four maxillary teeth underlying the orbit. HPG-15-3 has more teeth than any other known Mesozoic bird, which highlights both the importance of dentition in the feeding strategies of Cretaceous avians and the existence of strong positive selection for teeth despite the overall avian trend towards edentulism²⁹. Increased dentition in Cretaceous birds has previously been correlated with piscivory²⁹. However, the papillae preserved on the tongue and ventral surface of the upper jaw of *Oculudentavis* (Extended Data Fig. 2a, b)

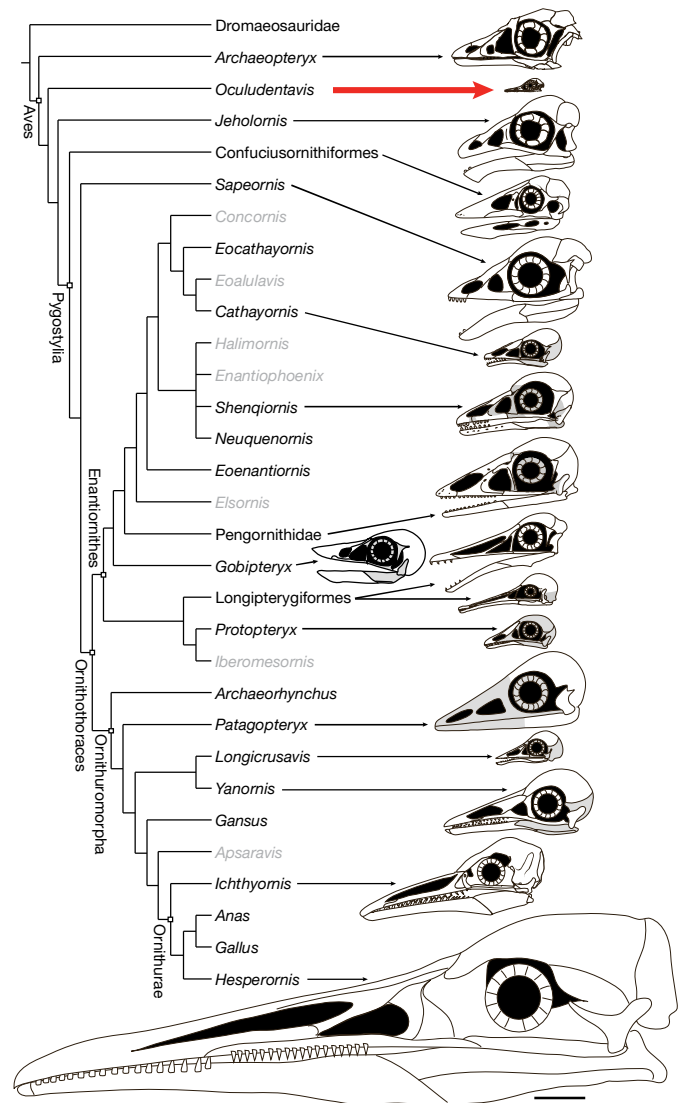


Fig. 4 | Simplified results of the strict consensus of 2,044 trees depicting the phylogenetic relationships of *O. khaungraae* relative to other known Mesozoic birds. The analysis (consisting of 36 taxa scored across 257 characters) was run in the TNT program using implied weighting ($k = 16$). Consistency index = 0.47; retention index = 0.645. In the analysis using a priori weights, *Oculudentavis* was resolved as an enantiornithine in a small percentage of the most-parsimonious trees. For most species for which skull material is known, the skull is reconstructed to scale relative to *O. khaungraae*. Scale bar, 2 cm. Names of species for which no skull material is known are written in grey.

lack the long, sharp and caudally oriented morphology that is present in many extant piscivorous birds¹⁴. Although unexpected for its small size, the large number of teeth—as well as their sharp and carinated morphology—suggests that *Oculudentavis* was a predator, probably feeding on small arthropods and other invertebrates (which are abundant in amber inclusions from the Angbamo site). The evolution of island nanism has been linked to a previously unrecognized ecological guild specialized in feeding on small invertebrates in *Paedophryne* frogs (a group that includes the smallest known vertebrate in the world³⁰), and a similar diet is possible for *Oculudentavis*.

Oculudentavis retains the plesiomorphic diapsid condition that is present in most nonornithothoracine birds, whereas the infratemporal bar was independently lost among enantiornithines and ornithuromorphs²⁷. Retention of the postorbital bar suggests limited cranial kinesis¹⁹, consistent with the absence of a distinct craniofacial hinge and with data that suggest avian cranial kinesis is mostly a morphological novelty of neognaths³¹.

The largely akinetic skull morphology of *Oculudentavis* suggests a bite force that is relatively stronger than that of a neornithine of the same size³², which is supported by the unusual degree of cranial fusion³³, the presence of a well-developed coronoid process and the absence of an antorbital fenestra. A distinct coronoid process is not developed in any other Mesozoic bird but it is well-developed in predatory birds of the Cenozoic era, such as phorusrhacids³⁴. In archosaurs, this process provides attachment for adductor muscles³⁵. Reduction in pneumatic features may suggest that, in the *Oculudentavis* lineage, solid bone was favoured biomechanically over opportunistic pneumatization³⁶. This reinforcement of the skull may reflect the predatory lifestyle inferred for *Oculudentavis*. Alternatively, the loss of the antorbital fenestra may be a consequence of miniaturization, and provide additional space for the enlarged orbit⁹.

HPG-15-3 preserves no features that unambiguously indicate its phylogenetic position, although the sum of the observed morphologies suggests that the taxon falls outside Ornithuromorpha. The extensive tooth row resembles noncoelurosaur theropods²⁶ and the ornithurine *Ichthyornis*¹², whereas the presence of premaxillary teeth (which are absent in the ornithurines *Hesperornis* and *Ichthyornis*), complete infratemporal and supratemporal fenestrae (which are typically absent in Ornithothoraces)³⁷, the quadrate with a broad orbital process (which is typically narrow in Ornithuromorpha) and the absence of a predentary (which is present in toothed Cretaceous ornithuromorphs)³⁸ all suggest a more-basal position. However, nonornithothoracine avian lineages (for example, *Jeholornis* and *Sapeornis*) are typically much larger than Cretaceous ornithothoracines—particularly the enantiornithines. Miniaturization is strongly associated with extreme homoplasy, which obscures phylogeny⁹. Our cladistic analysis using a priori weights produces a poorly resolved consensus tree in which all birds more derived than *Archaeopteryx* form a polytomy (Extended Data Fig. 10a). Approximately 90% of all equal-length trees resolve *Oculudentavis* as more derived than *Archaeopteryx* but basal to *Jeholornis*; the remaining 10% place it within Enantiornithes. Using implied weights (which is recommended for taxa characterized by strong homoplasy, including Aves^{39–41}), our analysis consistently resolves *Oculudentavis* as intermediate between *Archaeopteryx* and *Jeholornis* (Fig. 4, Extended Data Fig. 10b, Supplementary Information). If this hypothesis were to be supported by additional postcranial data, HPG-15-3 could possibly represent the first nonenantiornithine bird in the avifauna in the Hukawng valley deposits. Given the unusual morphology of *Oculudentavis* (which is clearly unlike that of any other bird), there is a strong potential for new data to markedly alter our systematic conclusion. Regardless, this discovery highlights the capacity of amber to document vertebrate diversity that is unrecorded by other depositional environments.

Note added in proof: Since this paper was accepted, an eight specimen that preserves the skeletal remains of an enantiornithine bird has been described⁴².

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2068-4>.

- Xing, L. et al. Mummified precocial bird wings in mid-Cretaceous Burmese amber. *Nat. Commun.* **7**, 12089 (2016).
- Xing, L. et al. A feathered dinosaur tail with primitive plumage trapped in mid-Cretaceous amber. *Curr. Biol.* **26**, 3352–3360 (2016).
- Daza, J. D. et al. An enigmatic miniaturized and attenuate whole lizard from the Mid-Cretaceous amber of Myanmar. *Breviora* **563**, 1–18 (2018).
- Xing, L.-D. et al. A mid-Cretaceous enantiornithine (Aves) hatchling preserved in Burmese amber with unusual plumage. *Gondwana Res.* **49**, 264–277 (2017).
- Xing, L.-D. et al. A flattened enantiornithine in mid-Cretaceous Burmese amber: morphology and preservation. *Sci. Bull. (Beijing)* **63**, 235–243 (2018).
- Xing, L. et al. A fully feathered enantiornithine foot and wing fragment preserved in mid-Cretaceous Burmese amber. *Sci. Rep.* **9**, 927 (2019).

- Xing, L., McKellar, R. C., O'Connor, J. K., Niu, K. & Mai, H. A mid-Cretaceous enantiornithine foot and tail feather preserved in Burmese amber. *Sci. Rep.* **9**, 15513 (2019).
- Xing, L. et al. A new enantiornithine bird with unusual pedal proportions found in amber. *Curr. Biol.* **29**, 2396–2401.e2 (2019).
- Hanken, J. & Wake, D. B. Miniaturization of body size: organismal consequences and evolutionary significance. *Annu. Rev. Ecol. Syst.* **24**, 501–519 (1993).
- Westerweel, J. et al. Burma Terrane part of the Trans-Tethyan Arc during collision with India according to palaeomagnetic data. *Nat. Geosci.* **12**, 863–868 (2019).
- Shi, G. et al. Age constraint on Burmese amber based on U-Pb dating of zircons. *Cretac. Res.* **37**, 155–163 (2012).
- Field, D. J. et al. Complete *Ichthyornis* skull illuminates mosaic assembly of the avian head. *Nature* **557**, 96–100 (2018).
- Smith, R. D. A. & Ross, A. Amberground pholidid bivalve borings and inclusions in Burmese amber: implications for proximity of resin-producing forests to brackish waters, and the age of the amber. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **107**, 239–247 (2018).
- Lovette, I. J. & Fitzpatrick, J. W. *The Handbook of Bird Biology* 3rd edn (Princeton Univ. Press, 2004).
- Dalsgaard, B. et al. Trait evolution, resource specialization and vulnerability to plant extinctions among Antillean hummingbirds. *Proc. R. Soc. Lond. B* **285**, 20172754 (2018).
- Glaw, F., Köhler, J., Townsend, T. M. & Vences, M. Rivaling the world's smallest reptiles: discovery of miniaturized and microendemic new species of leaf chameleons (*Brookesia*) from northern Madagascar. *PLoS ONE* **7**, e31314 (2012).
- Yeh, J. The effect of miniaturized body size on skeletal morphology in frogs. *Evolution* **56**, 628–641 (2002).
- Griffith, H. Miniaturization and elongation in *Eumeces* (Sauria: Scincidae). *Copeia* **1990**, 751–758 (1990).
- Chiappe, L. M., Ji, S., Ji, Q. & Norell, M. A. Anatomy and systematics of the Confuciusornithidae (Theropoda: Aves) from the Late Mesozoic of northeastern China. *Bull. Am. Mus. Nat. Hist.* **242**, 1–89 (1999).
- Elzanowski, A. Embryonic bird skeletons from the Late Cretaceous of Mongolia. *Palaeontologica Polonica* **42**, 147–179 (1981).
- Jollie, M. T. The head skeleton of the chicken and remarks on the anatomy of this region in other birds. *J. Morphol.* **100**, 389–436 (1957).
- Edinger, T. *Über Knöchernen Scleralringe* (Fisher, 1929).
- Schmitz, L. Quantitative estimates of visual performance features in fossil birds. *J. Morphol.* **270**, 759–773 (2009).
- Schmitz, L. & Motani, R. Morphological differences between the eyeballs of nocturnal and diurnal amniotes revisited from optical perspectives of visual environments. *Vision Res.* **50**, 936–946 (2010).
- Schmitz, L. & Motani, R. Nocturnality in dinosaurs inferred from scleral ring and orbit morphology. *Science* **332**, 705–708 (2011).
- Rauhut, O. W. M. *The Interrelationships and Evolution of Basal Theropod Dinosaurs (Special Papers in Palaeontology 69)* (The Palaeontological Association, London, 2003).
- O'Connor, J. & Chiappe, L. M. A revision of enantiornithine (Aves: Ornithothoraces) skull morphology. *J. Syst. Palaeontology* **9**, 135–157 (2011).
- Xu, X. & Norell, M. A. A new troodontid dinosaur from China with avian-like sleeping posture. *Nature* **431**, 838–841 (2004).
- O'Connor, J. K. The trophic habits of early birds. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **513**, 178–195 (2019).
- Rittmeyer, E. N., Allison, A., Gründler, M. C., Thompson, D. K. & Austin, C. C. Ecological guild evolution and the discovery of the world's smallest vertebrate. *PLoS ONE* **7**, e29797 (2012).
- Hu, H. et al. Evolution of the vomer and its implications for cranial kinesis in Paraves. *Proc. Natl Acad. Sci. USA* **116**, 19571–19578 (2019).
- Bout, R. G. & Zveers, G. A. The role of cranial kinesis in birds. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **131**, 197–205 (2001).
- Rafield, E. J. Aspects of comparative cranial mechanics in the theropod dinosaurs *Coelophysis*, *Allosaurus* and *Tyrannosaurus*. *Zool. J. Linn. Soc.* **144**, 309–316 (2005).
- Degrange, F. J., Tambussi, C. P., Taglioretti, M. L., Dondas, A. & Scaglia, F. A new Mesembriornithinae (Aves, Phorusrhacidae) provides new insights into the phylogeny and sensory capabilities of terror birds. *J. Vertebr. Paleontol.* **35**, e912656 (2015).
- Holliday, C. M. & Witmer, L. M. Archosaur adductor chamber evolution: integration of musculoskeletal and topological criteria in jaw muscle homology. *J. Morphol.* **268**, 457–484 (2007).
- Witmer, L. M. The evolution of the antorbital cavity of archosaurs: a study in soft-tissue reconstruction in the fossil record with an analysis of the function of pneumaticity. *J. Vertebr. Paleontol.* **17**, 1–73 (1997).
- O'Connor, J. K., Chiappe, L. M. & Bell, A. in *Living Dinosaurs: the Evolutionary History of Birds* (eds Dyke, G. D. & Kaiser, G.) 39–114 (John Wiley & Sons, 2011).
- Bailleul, A. M., Li, Z., O'Connor, J. & Zhou, Z. Origin of the avian predentary and evidence of a unique form of cranial kinesis in Cretaceous ornithuromorphs. *Proc. Natl Acad. Sci. USA* **116**, 24696–24706 (2019).
- Zhou, Z. & Zhang, F. A long-tailed, seed-eating bird from the Early Cretaceous of China. *Nature* **418**, 405–409 (2002).
- Xu, X. Mosaic evolution in birds: brain vs. feeding apparatus. *Sci. Bull. (Beijing)* **63**, 812–813 (2018).
- Goloboff, P. A., Carpenter, J. M., Arias, J. S. & Esquivel, D. R. M. Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* **24**, 758–773 (2008).
- Xing, L.-D., McKellar, R. C. & O'Connor, J. An unusually large bird wing in mid-Cretaceous Burmese amber. *Cretaceous Res.* **110**, 104412 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Material and photography

HPG-15-3 is housed and displayed in the Hupoge Amber Museum; the 3D printing models are available through the Dexu Institute of Palaeontology. HPG-15-3 was examined with a VH-Z100UT Universal Zoom Lens (100× to 1,000×) in an optical microscope (KEYENCE VHX-6000) with 100× magnification at Shenyang Normal University. Photographs were taken using a Canon digital camera (5D Mark III, MP-E 65MMF/2.81-5X) fitted to a macro rail (Cognisys) and processed using Helicon Focus 5.1 and Adobe Photoshop CS5 software.

Scanning and 3D reconstruction

HPG-15-3 was imaged nondestructively using propagation phase-contrast synchrotron radiation X-ray microtomography, on beamline 13W of the Shanghai Synchrotron Radiation Facility. The SR beam was monochromatized at 22 keV using the double Si (111) crystal monochromator. The distance between sample and detector (propagation distance) was 60 mm to obtain the phase contrast. The physical pixel size of the charge-coupled device sensor was $6.5 \mu\text{m} \times 6.5 \mu\text{m}$, and we used a 2× microscope objective; the isotropic voxel size was $3.25 \mu\text{m}$. The pixel number of our detector was $2,048 \times 2,048$ and its dynamic range was 16 bit. The field of the view of our detector was $6.6 \text{ mm} \times 6.6 \text{ mm}$.

Six microcomputed tomography scans under the same conditions were performed to image the whole specimen, owing to the fact its size in the vertical direction is much larger than the field of view of our detector. The off-axis scanning 360° mode was adopted to 3 scans because the size of the sample in the horizontal direction was larger than the field of view of the detector. In this mode, 6,000 projections were collected for a single computed tomography scan. The exposure time of a single projection was 0.3 s.

In the preprocessing of the off-axis scanning data, image stitching was the first step performed, considering the image alignment, the contrast unification and the tilt of the rotation axis. We obtained 3,000 stitched images with enlarged horizontal fields of view (about $6.6 \text{ mm} \times 13 \text{ mm}$) for 3,000 project angles. The phase retrieval and slice reconstruction of all the six projection datasets were performed using PITRE-3 software⁴³. After the reconstruction, the computed-tomography slice sets for the six segments were stitched again in the vertical direction.

The computed tomography slices for the entire sample comprised 36 GB of data. To reduce the amount of memory used, the sampling interval was set as one pixel in all three directions, so that the data were reduced to one-eighth of their original size. The amber and all

impurities surrounding the skull were removed using the image segmentation function. The rotation correction was performed using ImageJ software. The 3D data processing, segmentation and analysis were performed using VG StudioMax 1.2 and 2.1. The 3D morphology of the endocast was obtained through manual segmentation using ImageJ 1.4 and rendered using VG studio Max 2.4.

Scaling of eye socket and skull length

We assessed the allometry of the avian eye socket with phylogenetic generalized least squares⁴⁴. We used a Brownian correlation matrix and maximum likelihood estimates of λ to summarize the slope estimates over 1,000 previously published time-calibrated trees^{45,46}.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Owing to their size, the raw computed tomography data are available upon request from L.X. (xinglida@gmail.com). All other materials are included in the Supplementary Information or are available at <https://doi.org/10.5281/zenodo.3591994>.

43. Chen, R.-C. et al. PITRE: software for phase-sensitive X-ray image processing and tomography reconstruction *J. Synchrotron Radiat.* **19**, 836–845 (2012).
44. Symonds, M. R. E. & Blomberg, S. P. in *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology* (ed. Garamszegi, L. Z.) 105–130 (Springer, 2014).
45. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 309–316 (2012).
46. Jetz, W. et al. Distribution and conservation of global evolutionary distinctness in birds. *Curr. Biol.* **24**, 919–930 (2014).

Acknowledgements This research was funded by the National Natural Science Foundation of China (no. 41888101, 41790455 and 41772008), the National Geographic Society (no. EC0768-15) and the Natural Sciences and Engineering Research Council of Canada (2015-00681). We thank BL13W of the Shanghai Synchrotron Radiation Facility for beamtime access based on proposal 16ssrf 01737, and the Beijing Synchrotron Radiation Facility for supplying the high MTF imaging detector. We thank S. Abramowicz for assistance with figures and D. Blackburn, D. Steadman and E. Stanley for making the computed tomography scan of *M. minima* accessible.

Author contributions L.X. and J.K.O. designed the project, L.X., J.K.O., L.M.C., L.S., R.C.M., Q.Y. and G.L. performed the research: G.L. and Q.Y. performed computed tomography scanning of the specimen and processed the data. L.S. performed the eye-scaling statistical analyses. J.K.O. performed the cladistic analysis. J.K.O., L.M.C., L.S., L.X. and G.L. wrote the manuscript. L.X., J.K.O., L.S. and G.L. contributed equally.

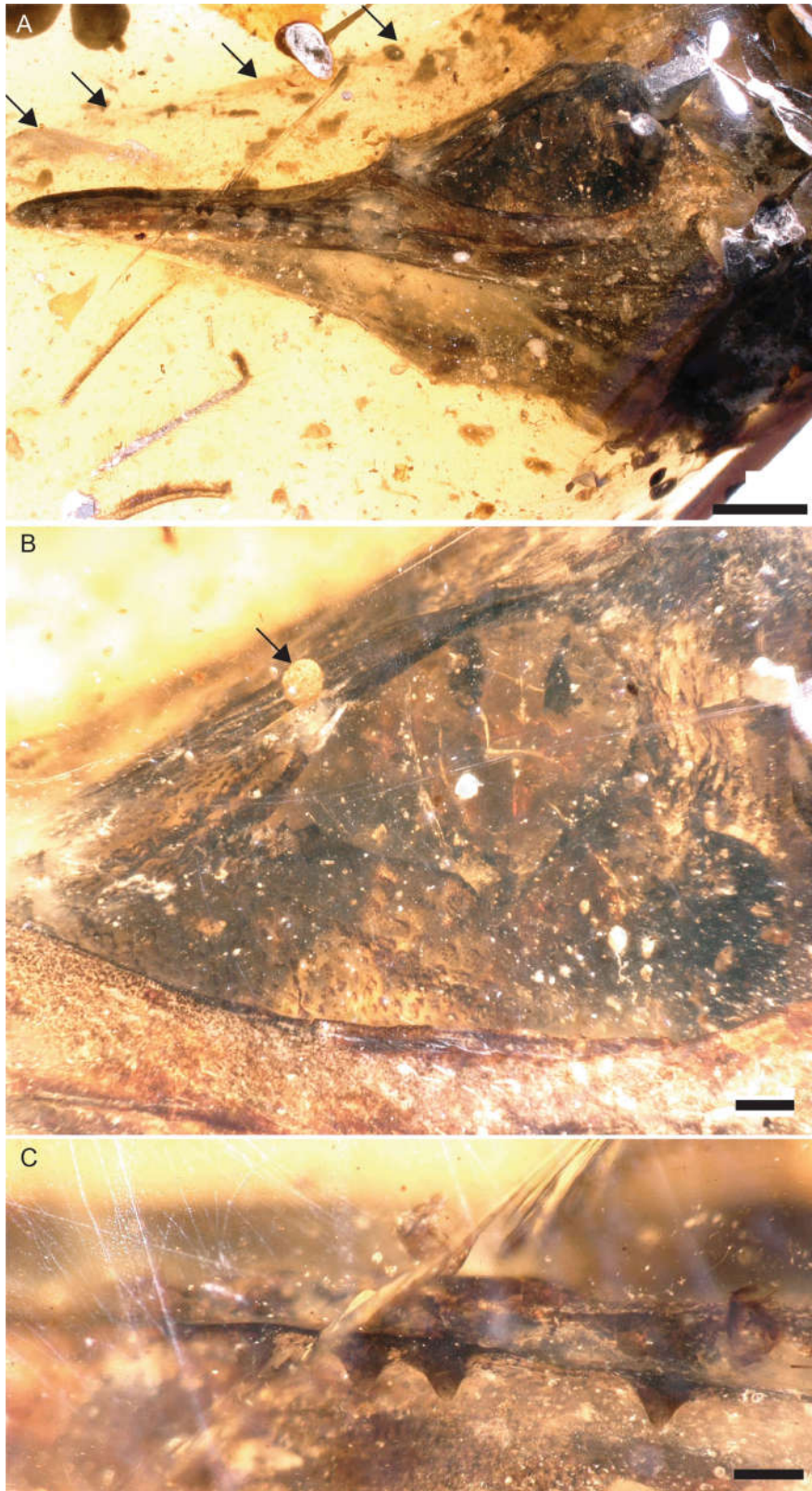
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2068-4>.

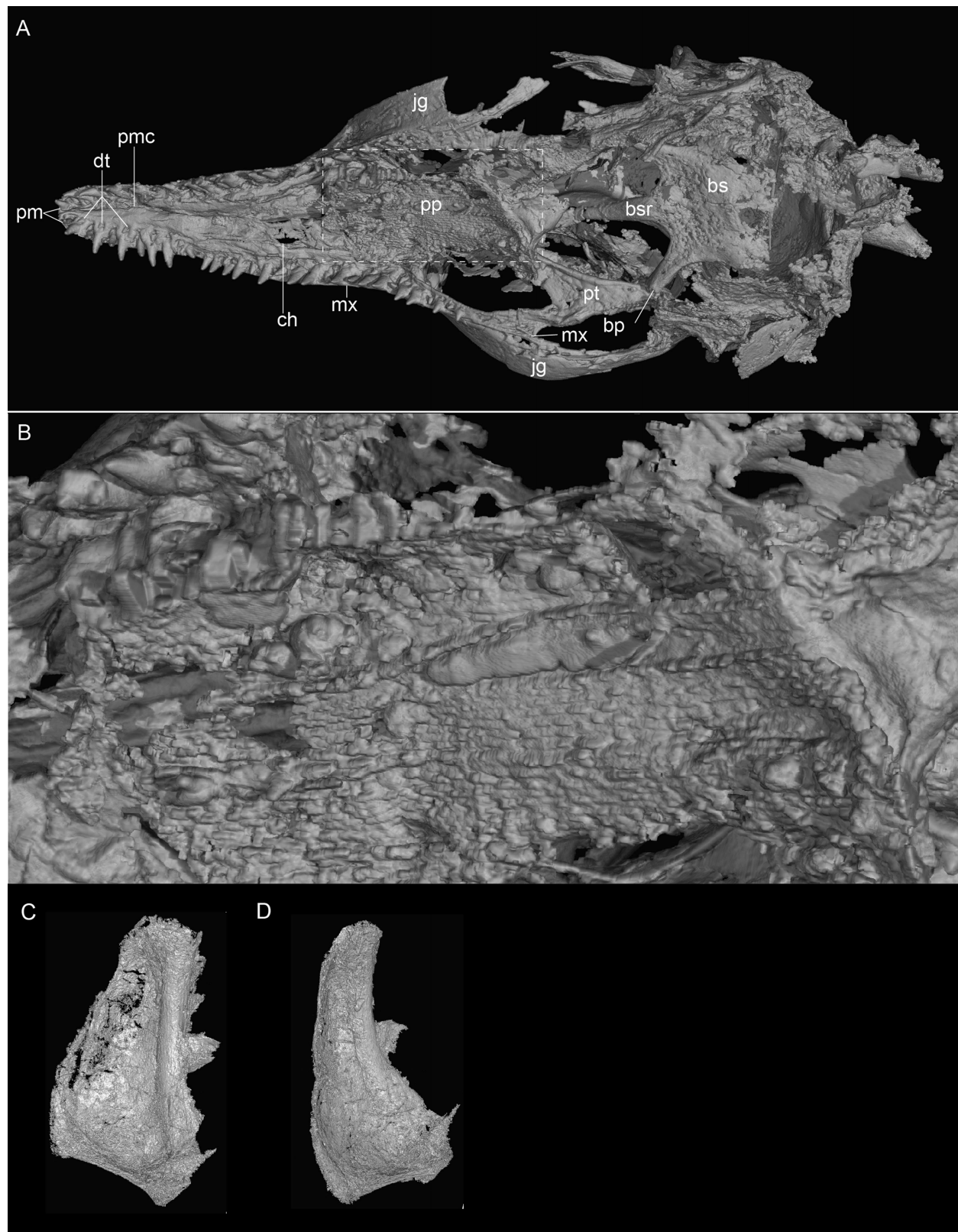
Correspondence and requests for materials should be addressed to J.K.O.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



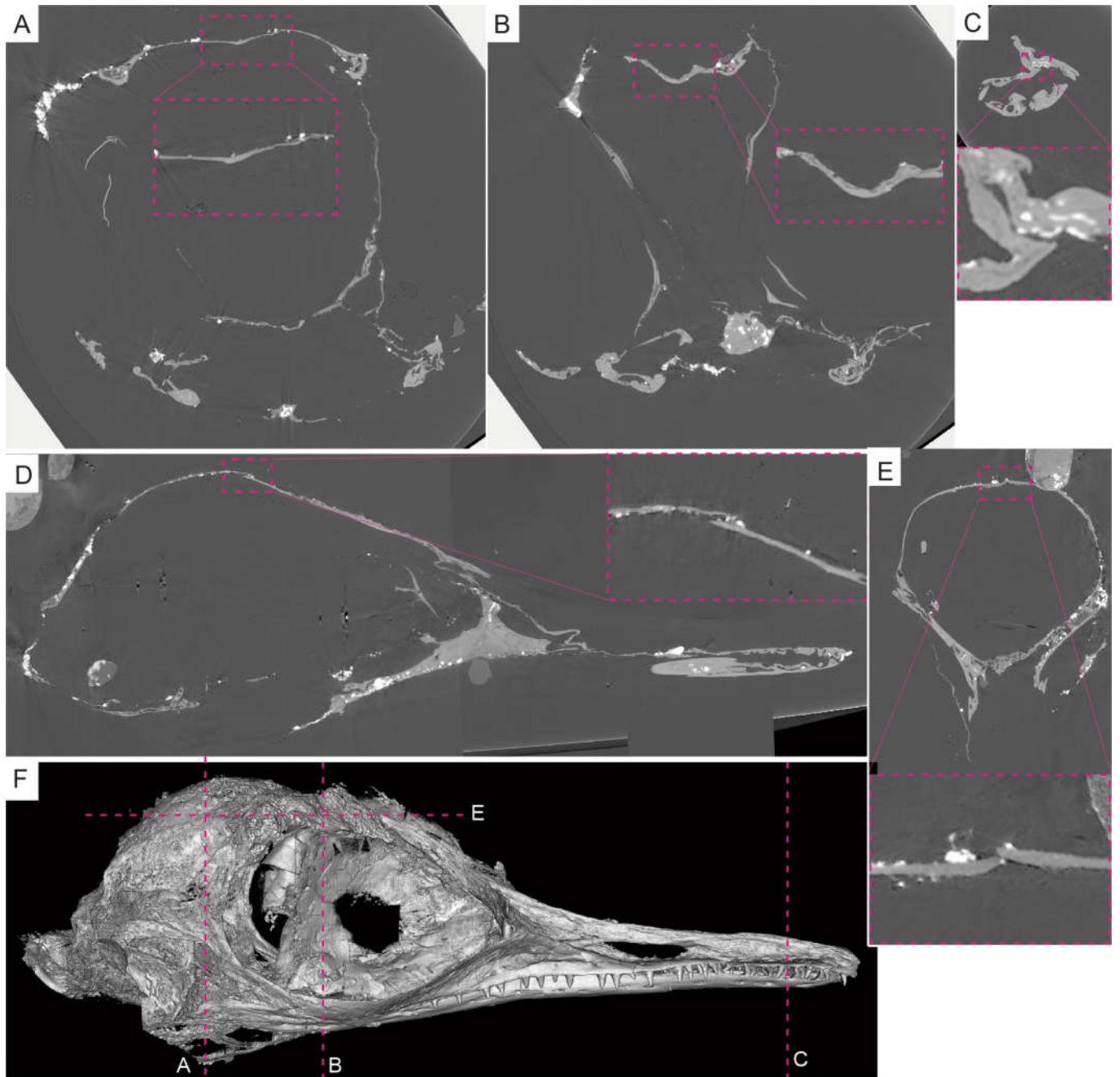
Extended Data Fig. 1 | Close-up photographs of HPG-15-3. a, Entire skull in left lateral view. **b,** Left eye. **c,** Maxillary dentition. The black arrows in **a** indicate decay products from the soft tissue of the dorsal surface of the skull and the

original position of skull, which drifted before the resin hardened; the black arrow in **b** indicates the position of decay products released from the left eye. Scale bars, 2 mm (**a**), 500 μ m (**b**), 200 μ m (**c**).



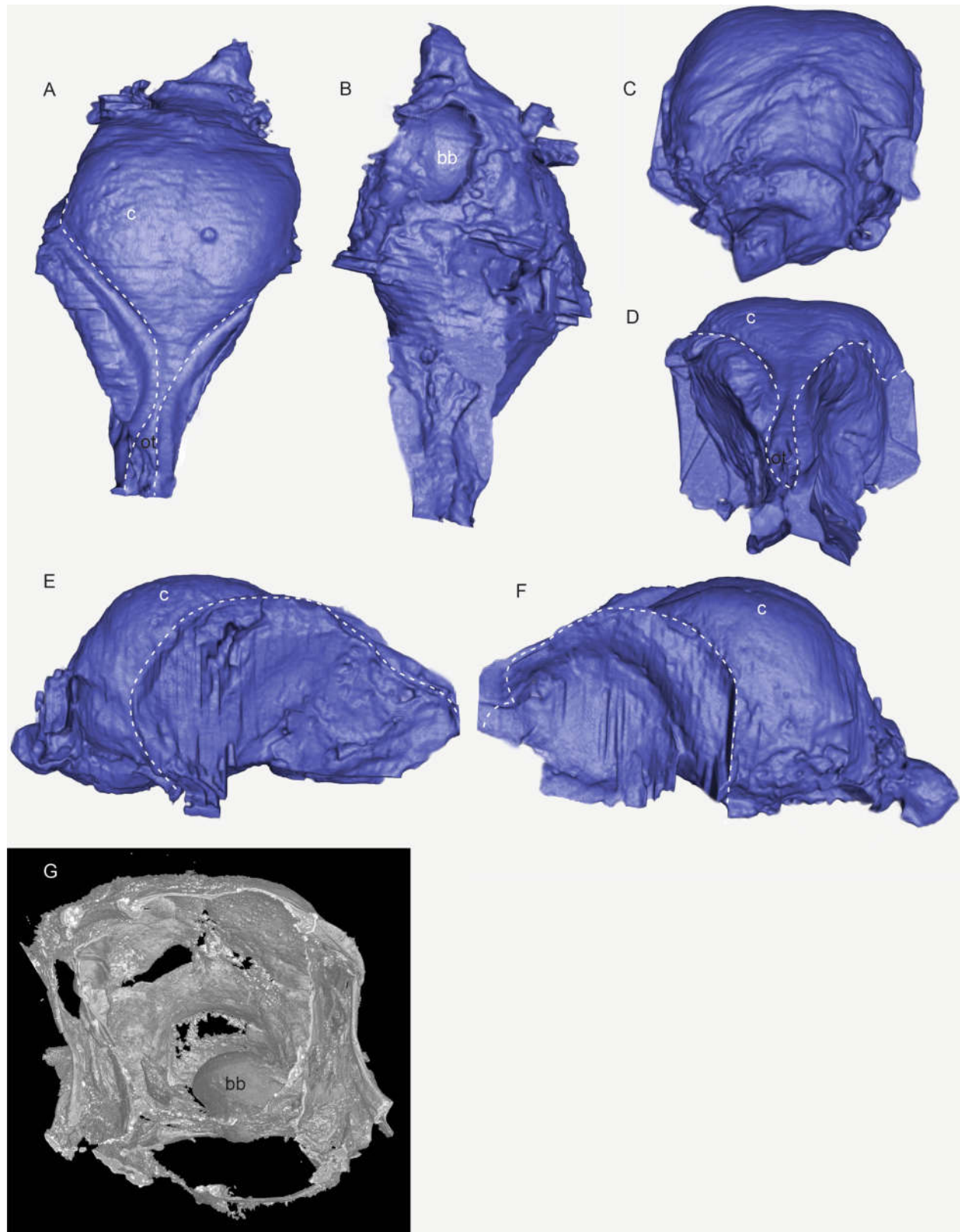
Extended Data Fig. 2 | Computed tomography scan of HPG-15-3 in palatal view, with the mandibles removed, and an isolated quadrate. a, Full palatal view. **b,** Close-up of the preserved lingual papillae of the roof of the mouth. **c,** Isolated left quadrate in lateral view. **d,** Quadrate in caudal view. Dashed

square box in **a** indicates the region enlarged in **b**. bp, basipterygoid process; bs, basisphenoid plate; bsr, basisphenoid rostrum; ch, choana; dt, developing tooth; pt, pterygoid; pp, papillae; pmc, medial contact of the palatal processes of the premaxillae.



Extended Data Fig. 3 | Raw computed tomography slices showing the anatomy of important cranial sutures of HPG-15-3. a, Interparietal suture, cranial portion (closed). **b,** Interfrontal suture (closed). **c,** Palatal processes of the premaxilla (open). **d,** Frontoparietal suture (open). **e,** Interparietal suture,

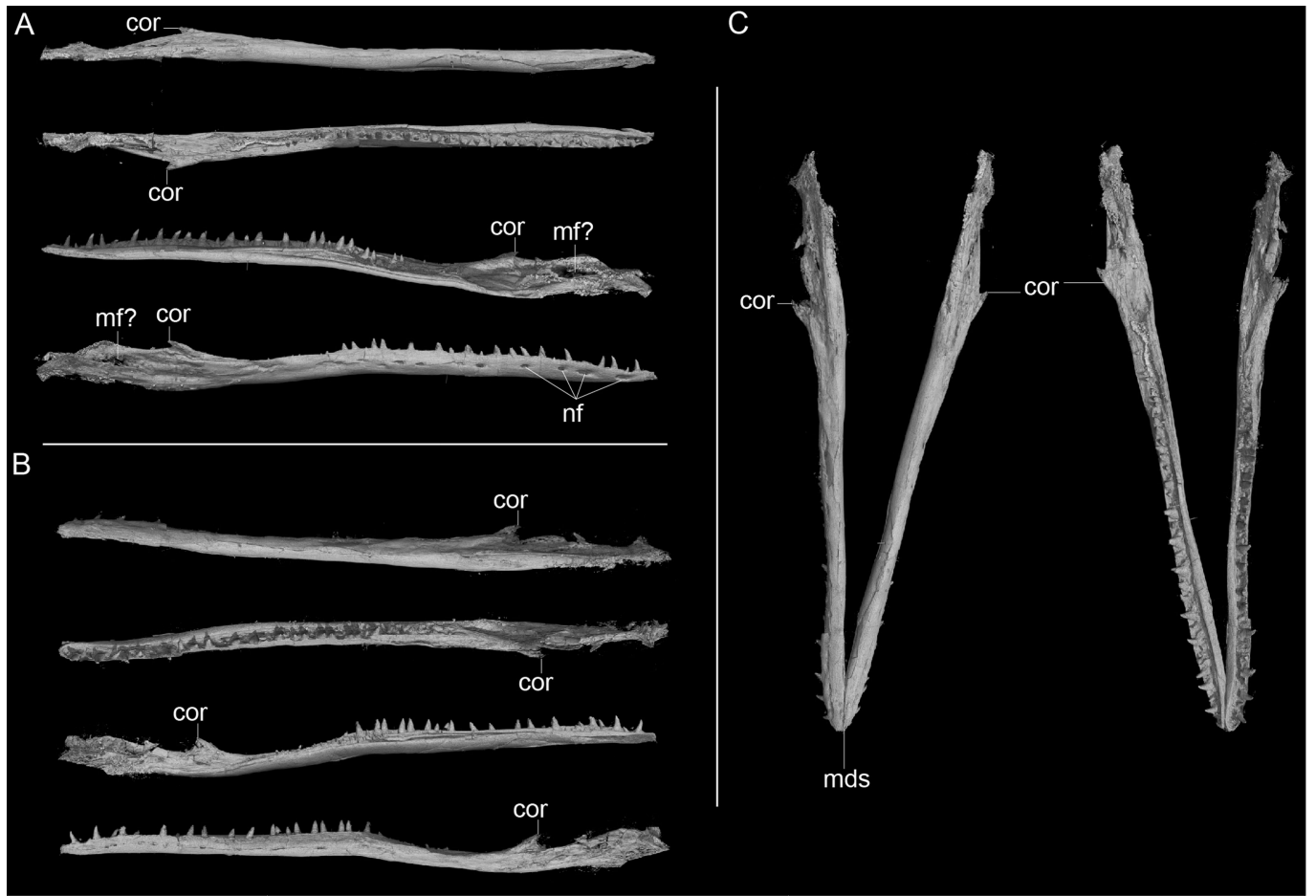
caudal portion (open). **f,** Image of the entire skull, showing position of the slices shown in **a–c, e** (image shown in **d** is a sagittal slice through the middle of the skull). Boxes outlined in dashed pink lines show the region enlarged in the insets, to clearly demonstrate the morphology of the suture or contact.



Extended Data Fig. 4 | Rendering of the cranial endocast of HPG-15-3.

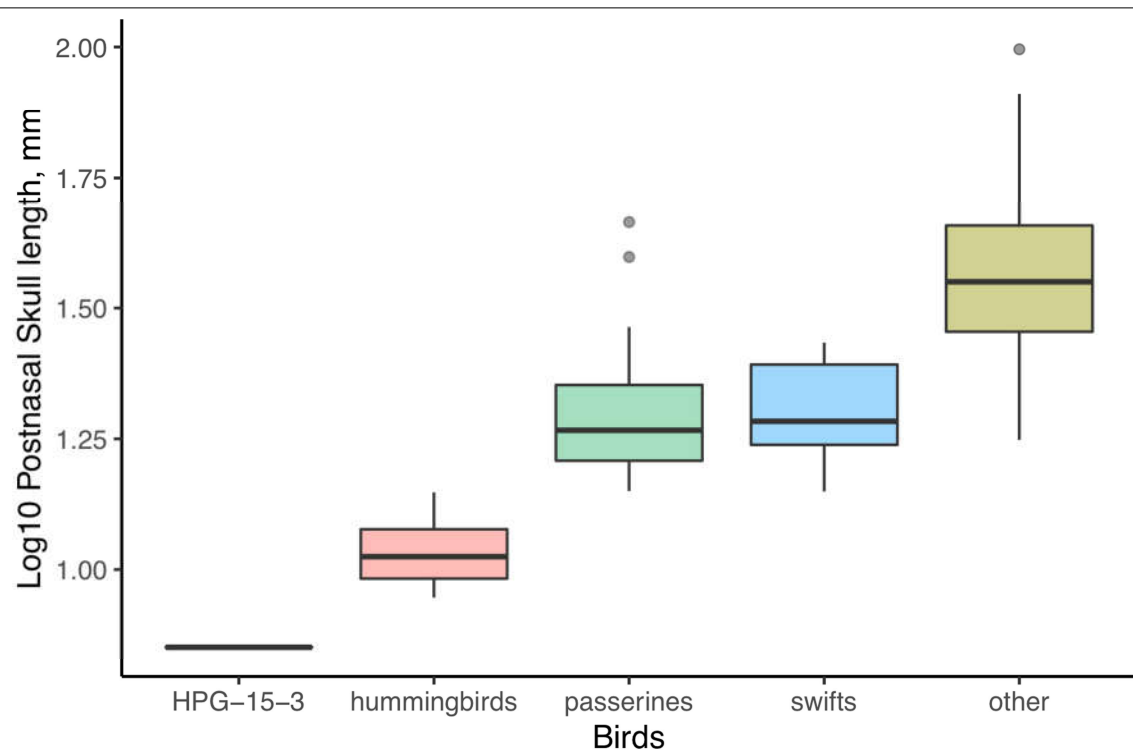
a, Dorsal view. **b**, Ventral view. **c**, Caudal view. **d**, Cranial view. **e**, Right lateral view. **f**, Left lateral view. **g**, Interior view of the brain cavity, showing the bivale boring that intrudes through the ventral surface. Because the ventral surface of the cranium is damaged by a bivale boring and the bones supporting the cranial margins of the brain are not preserved, only the dorsal surface of the

endocast reveals reliable information. The white dashed lines indicate the portions of the endocast that probably were not occupied by brain tissue. The cerebrum appears to be prominent but a distinct optic lobe—as seen in other birds—cannot be identified. **bb**, bivale boring; **c**, cerebrum; **ot**, part of the olfactory tract and/or olfactory lobe.



Extended Data Fig. 5 | Isolated mandible of HPG-15-3. a, Right mandible in (from top to bottom) ventral, dorsal, medial and lateral views. **b,** Left mandible in (from top to bottom) ventral, dorsal, medial and lateral views. **c,** Articulated

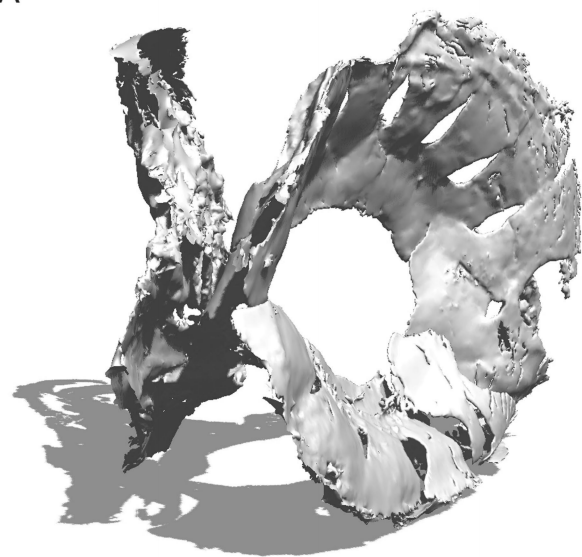
mandibles in ventral (left) and dorsal (right) views. cor, coronoid process; mds, mandibular symphysis; mf?, possible mandibular foramen; nf, nutrient foramina.



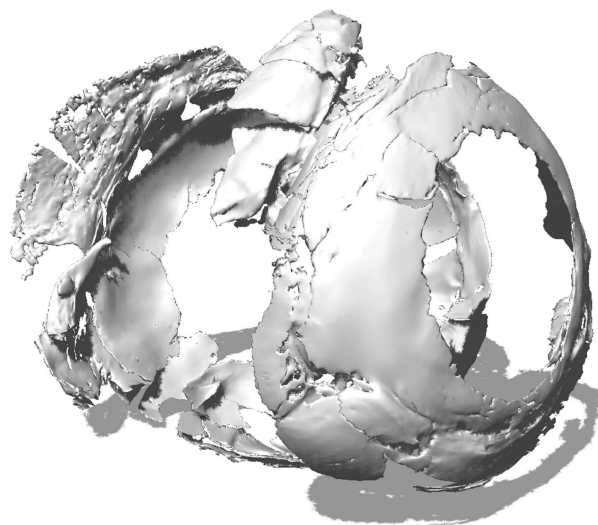
Extended Data Fig. 6 | Skull size in HPG-15-3 compared to other birds. The skull of HPG-15-3 is small compared to those of extant birds (total of $n = 213$ extant bird species sampled), here illustrated through a box plot of \log_{10} -transformed postnasal skull length (as a proxy for the braincase), measured from the craniofacial hinge to the caudal end of the cranium. Each box plot

illustrates the median (thick line) the 1st and 3rd quartiles (the hinges), and the distance from the upper and lower hinge to the largest and smallest value no further than $1.5 \times$ the interquartile range (the whiskers). HPG-15-3 is smaller than swifts ($n = 12$), passerines ($n = 23$) and hummingbirds ($n = 22$)—and may even be smaller than the smallest hummingbird, *M. helenae*.

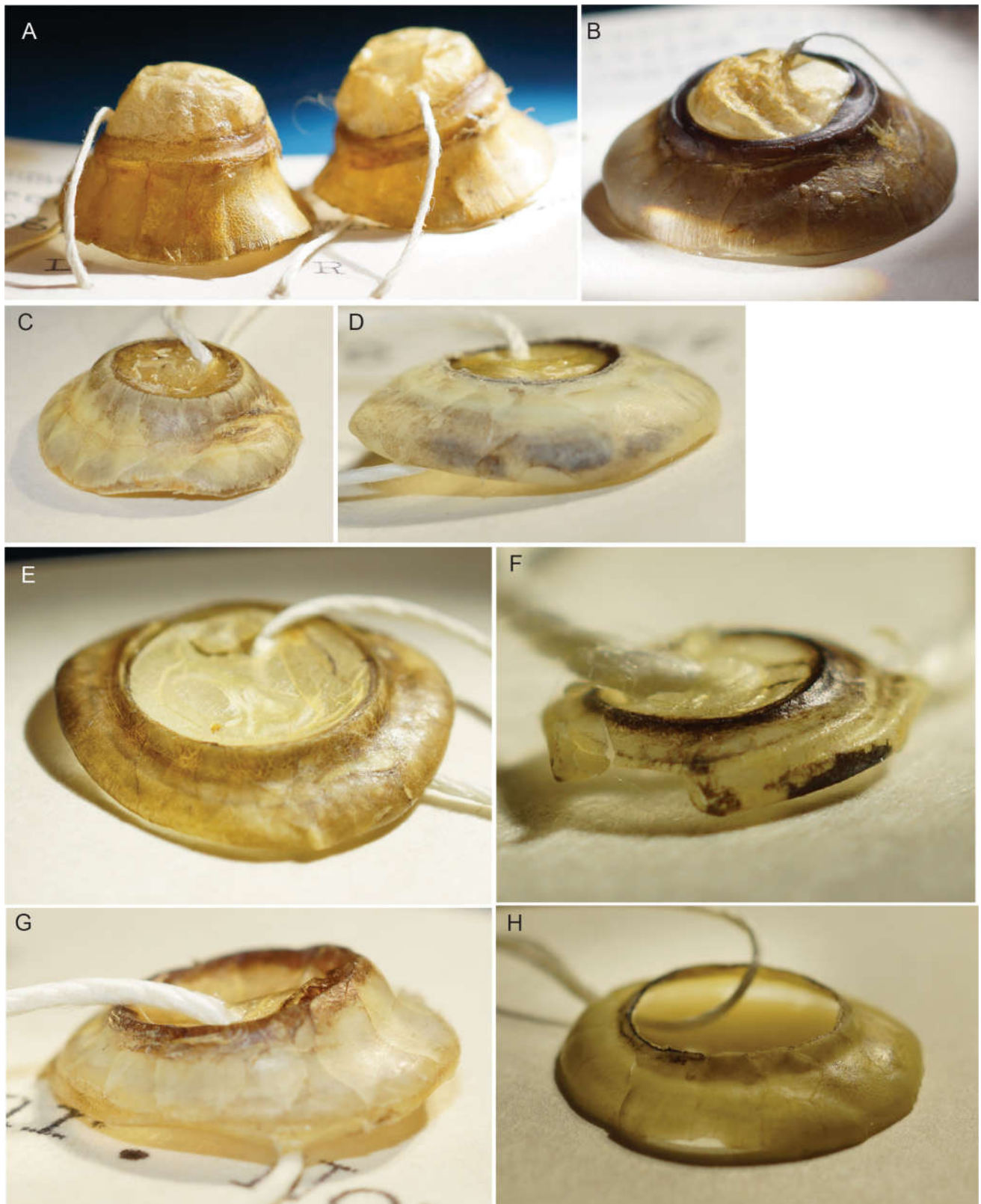
A



B



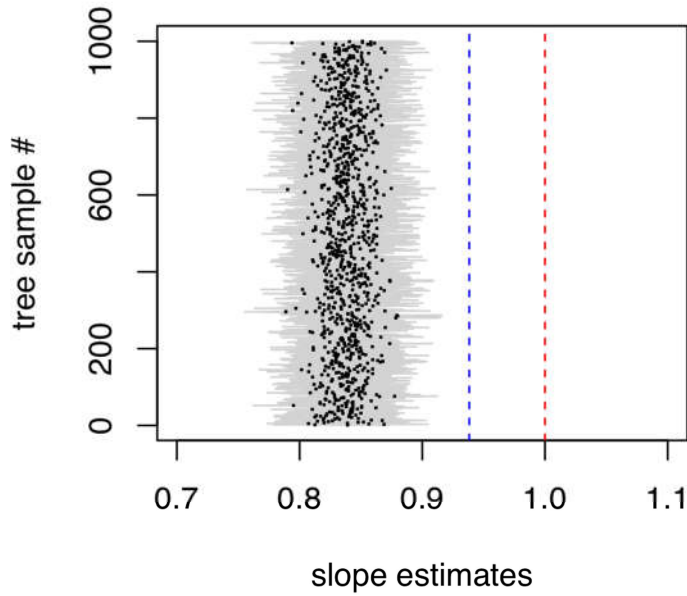
Extended Data Fig. 7 | Scleral ring of HPG-15-3. a, As preserved. b, After retrodeformation.



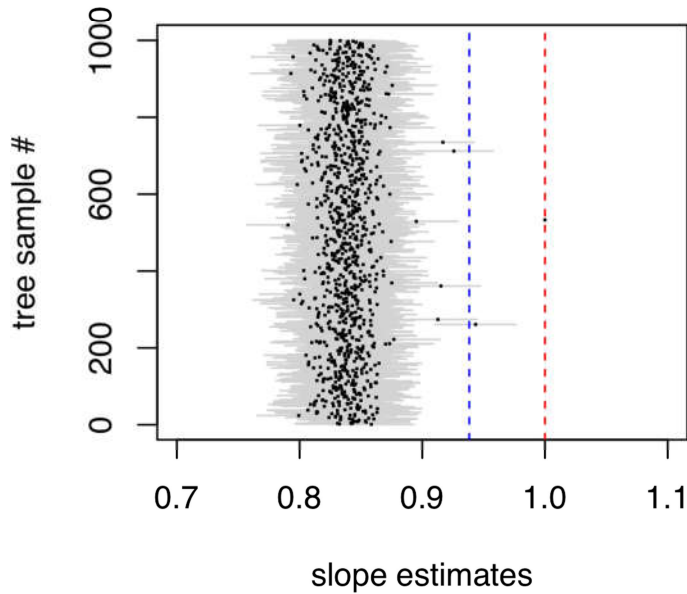
Extended Data Fig. 8 | Scleral rings of selected neornithines. **a.** *Asio flammeus*, short-eared owl. **b.** *Buteo jamaicensis*, red-tailed hawk. **c.** *Cerorhinca monocerata*, rhinoceros auklet. **d.** *Dendrocopus villosus*, hairy woodpecker.

e. *Chordeiles minor*, common nighthawk. **f.** *Selaphorus sasin*, Allen's hummingbird. **g.** *Cypseloides niger*, American black swift. **h.** *Megasceryle alcyon*, belted kingfisher.

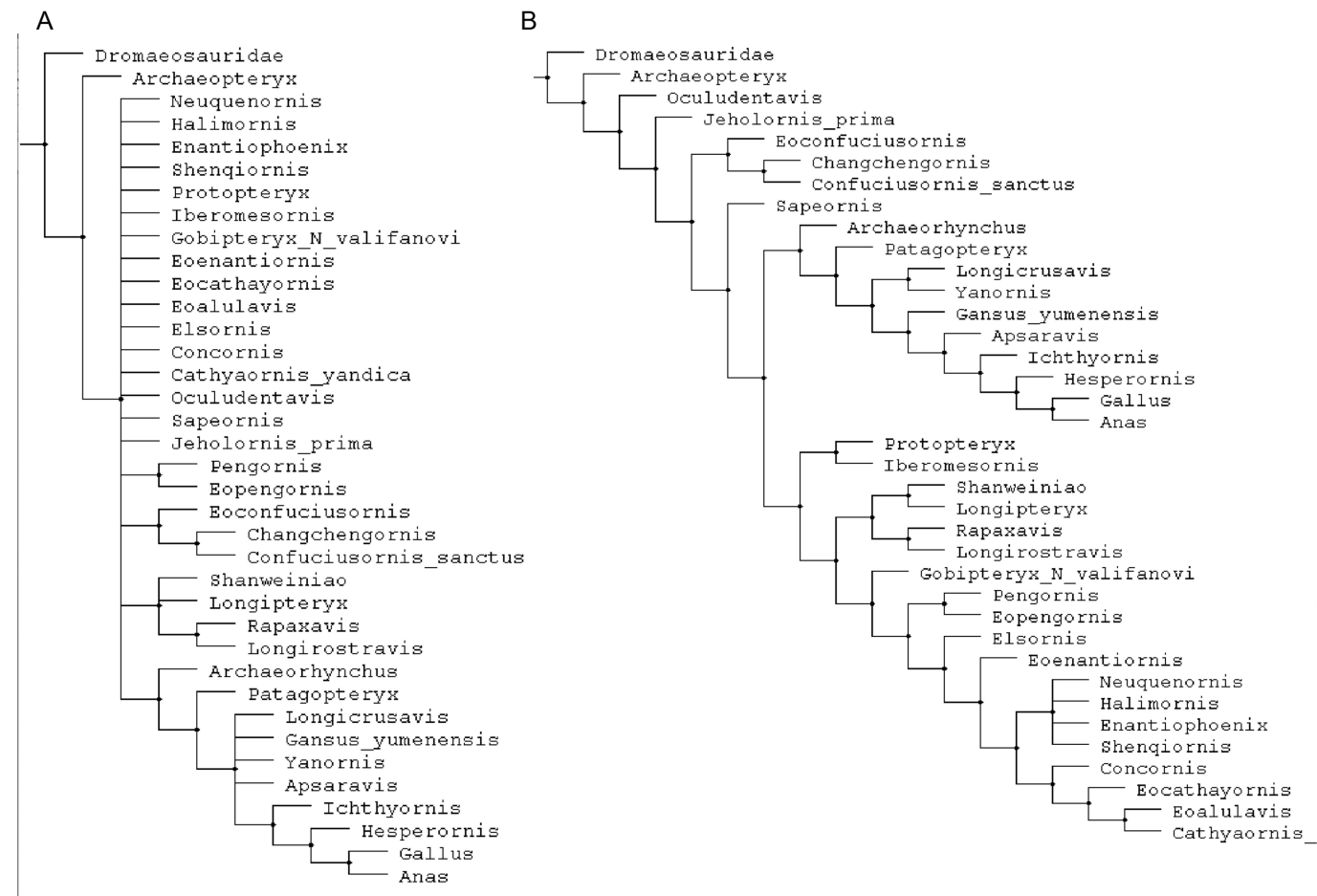
Slope estimates, Ericson trees



Slope estimates, Hackett trees



Extended Data Fig. 9 | Summary of the slope estimates obtained from phylogenetic generalized least squares with the Ericson and the Hackett backbone tree sets, containing 1,000 trees each. **a**, Ericson backbone tree set. **b**, Hackett backbone tree set. The *x* axes represent the estimated slope values, and the *y* axes represent the number of the tree sampled from the entire set of 1,000 trees. Dots signify the actual slope estimate, and grey bars visualize the s.e. of the slope estimates. The blue dashed line is the slope value obtained from ordinary least square regression, and the red dashed line represents the slope of isometry. Results from the phylogenetic generalized least squares iterations ($n=1,000$) suggest the presence of negative allometry (slope < 1) in the relation between eye socket and skull length ($P < 0.001$).



Extended Data Fig. 10 | Results of analysis using a priori weights and the unsimplified cladogram depicting the results of the phylogenetic analysis. For further discussion, see Supplementary Information. **a**, A priori weighting results in a polytomy consisting of all taxa more derived than *Archaeopteryx*. **b**, Results of the analysis using implied weighting with a k value of 16; results

were the same for k values of 12–20 and of 25; k values between 2 and 11 differed only from higher k values in the relative placement of some derived enantiornithines. Here *Nanantius valifanovi* is considered a junior synonym of *Gobipteryx minuta* (thus the operational taxonomic unit name is given as ‘*Gobipteryx_N_valifanovi*’).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used to collect data, except for the linear dimensions of the skull of <i>Mellisuga minima</i> : Lars Schmitz took measurements of a scaled digital model in Meshlab 2016.12.
Data analysis	All data were analysed in R 3.5.0 using routine phylogenetic comparative methods. Custom scripts are available in the Supplementary Information. Figures were made using Adobe CS 2018. Phylogenetic analysis was run using TNT 1.0. CT data was processed using ImageJ 1.4 and VG StudioMax 1.2, 2.1 and 2.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data are available in the Supplementary Information.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We describe the anatomy of a new fossil enclosed in Burmese amber, facilitated through synchrotron microtomography. We also performed large-scale phylogenetic comparative analysis of the scaling of eye socket size with respect to skull length, and visualization of eye shape through scleral ring and eye socket anatomy. Both approaches facilitate the comparison of the fossil with modern birds and lizards.
Research sample	The fossil we are describing in this paper is curated in the Hupoge Amber Museum, identified through the unique collection number HPG-15-3. To facilitate comparative analyses, data on eye socket and skull length were collected for a total of 206 extant species of birds. These samples were chosen to reflect all major clades of living birds, providing an overview of the overall pattern found across birds. Data on eye shape (orbit length, external and internal scleral ring diameter) and diel activity patterns of lizards and birds are from a previous study (Schmitz and Motani, 2011, Science).
Sampling strategy	Samples were chosen to reflect all major clades of living birds. Whenever possible, multiple individuals per species were measured; all analyses were performed with the average values of species. Regression analyses of smaller sample sizes (n=50, randomly chosen from the 206 species) revealed very similar results, hence the sample size used for the final analysis is deemed adequate.
Data collection	<p>Li Gang and and Yi Qiru conducted the CT scans and reconstruction - HPG-15-3 was imaged nondestructively using propagation phase contrast Synchrotron Radiation X-ray microtomography (PPC-SRX-μCT) on beamline 13W of the Shanghai Synchrotron Radiation Facility (SSRF). The SR beam was monochromatized at 22 keV using the double Si (111) crystal monochromator. The distance between sample and detector (propagation distance) was 60 mm to obtain the phase contrast. The physical pixel size of CCD sensor was 6.5μm×6.5μm and we used a 2? microscope objective; the isotropic voxel size was 3.25 μm. The pixel number of our detector was 2048 x 2048 and its dynamic range was 16 bit. The field of the view of our detector was 6.6 mm × 6.6 mm.</p> <p>6 μCT scans under the same conditions were performed to image the whole specimen due to the fact its size in the vertical direction is much larger than the FOV of our detector. The off-axis scanning 360° mode was adopted to 3 scans because the sample's size in horizontal direction was larger than detector's FOV. In this mode 6000 projections were collected for a single CT scan. The exposure time of single projection was 0.3 s.</p> <p>In the off-axis scanning data preprocessing, the image stitching was the first step that always required to consider the image alignment, the contrast unification, and the tilt of the rotation axis. We obtained 3000 stitched images with enlarged horizontal FOV, about 6.6 mm × 13 mm, for 3000 project angle. The phase retrieval and slice reconstruction of all the six projection data sets were performed using PITRE-3 software [S12]. After all the reconstruction, the six segments CT slice sets were stitched again in the vertical direction.</p> <p>The CT slices for the entire sample consisted of 36 GB data. To reduce the amount of memory used, the sampling interval was set as one pixel in all 3 directions so that the data were reduced to one-eighth of its original size. The amber and all impurities surrounding the skull were removed using the image segmentation function. The rotation correction was performed using ImageJ software. The 3-D data processing, segmentation, and analysis was performed using VG StudioMax 1.2 and 2.1. The 3d morphology of the endocast was obtained through manual segmentation using ImageJ 1.4 and rendered using VG studio Max 2.4.</p> <p>Jingmai O'Connor performed the phylogenetic analysis using TNT. HPG-15-3 was scored into an modified version of the matrix published by Atterholt et al. 2018 which in turn is a modified version of the 2013 matrix published by O'Connor and Zhou. In order to elucidate the phylogenetic position of HPG-15-3 we added five new cranial characters - O'Connor scored these characters from published data. The modified matrix consists of 36 birds scored across 257 morphological characters with Dromaeosauridae as the outgroup, which was analyzed using TNT (Goloboff et al 2008). First, a heuristic search was conducted using tree-bisection reconnection (TBR) retaining the single shortest tree from every 1,000 replications. This produced 46 trees with a length of 728 steps; an additional round of TBR produced 2,044 trees of the same length, producing a large polytomy in the Nelson Strict Consensus tree. Early avian evolution is extremely homoplastic thus we utilized implied weighting. We explored k values from one to 20 and 25 and found that the position of Oculudentavis stabilized at k values higher than one and that enantiornithine interrelationships stabilized at k values higher than 11 (enantiornithine relationships are also stable from k equals 2 to 11 and differ from higher k values only in the relative placement of a few deeply nested taxa). In the presented analysis we conducted a heuristic search using TBR, retaining the single shortest tree from every 1,000 replications with a k value of 16. This produced three most parsimonious trees with a score of 21.3. These trees differed only in the relative placement of the most derived enantiornithines, which form a small polytomy in the strict consensus tree (Consistency Index = 0.47; Retention Index = 0.645).</p> <p>Lars Schmitz collected data of the skull length and eye socket length of birds with digital calipers, in the collections of the Los Angeles County Museum of Natural History (LACM), California State University Long Beach (CSULB, now integrated with the LACM collections), and the Smithsonian National Museum of Natural History (USNM). Collection numbers of all specimens are provided in the data table. A cranium of <i>Mellisuga minima</i> (UF 19832) was CT scanned by Edward Stanley at the Nanoscale Research Facility at the University of Florida, Gainesville, and uploaded to Morphosource (identifier: S21840). CT acquisition data are provided at Morphosource (https://www.morphosource.org/Detail/MediaDetail/Show/media_id/39924).</p>

Timing and spatial scale	N/A
Data exclusions	No data were excluded.
Reproducibility	During the design phase of the study, linear dimensions were taken repeatedly for the same individual skeleton to ensure consistency. All repeated measurements were successful.
Randomization	Grouping of birds and lizards by diel activity pattern was taken from a previously published study.
Blinding	Blinding was not possible and also not relevant for this study. All bird specimens in museum collections are properly labeled hence it is impossible to not know what bird taxon is measured. Given that the main goal was to characterize the overall scaling of eye socket size, irrespective of taxon or ecological group, blinding was not deemed necessary.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Palaeontology

Specimen provenance	Angbamo locality, Hukawng Valley, Kachin Province, Myanmar; the source and preservation of HPG-15-3 are in full compliance with the laws of Myanmar and China, including Myanmar's import and export regulations of jewelry and China's laws and regulations on museum. All extant bird specimens that were included in this study are curated in the collections of the LACM (including the CSULB specimens), USNM, and the University of Florida.
Specimen deposition	HPG-15-3 is housed and displayed in the Hupoge Amber Museum (HPG), Tengchong City Amber Association, China.
Dating methods	No new dates are provided.
<input type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	

Recycling and metabolic flexibility dictate life in the lower oceanic crust

<https://doi.org/10.1038/s41586-020-2075-5>

Received: 16 August 2019

Accepted: 10 January 2020

Published online: 11 March 2020

 Check for updates

Jiangtao Li^{1,2,9}, Paraskevi Mara^{2,9}, Florence Schubotz^{3,4}, Jason B. Sylvan⁵, Gaëtan Burgaud⁶, Frieder Klein⁷, David Beaudoin², Shu Ying Wee⁵, Henry J. B. Dick², Sarah Lott², Rebecca Cox², Lara A. E. Meyer^{3,4}, Maxence Quémener⁶, Donna K. Blackman⁸ & Virginia P. Edgcomb^{2,9}✉

The lithified lower oceanic crust is one of Earth's last biological frontiers as it is difficult to access. It is challenging for microbiota that live in marine subsurface sediments or igneous basement to obtain sufficient carbon resources and energy to support growth^{1–3} or to meet basal power requirements⁴ during periods of resource scarcity. Here we show how limited and unpredictable sources of carbon and energy dictate survival strategies used by low-biomass microbial communities that live 10–750 m below the seafloor at Atlantis Bank, Indian Ocean, where Earth's lower crust is exposed at the seafloor. Assays of enzyme activities, lipid biomarkers, marker genes and microscopy indicate heterogeneously distributed and viable biomass with ultralow cell densities (fewer than 2,000 cells per cm³). Expression of genes involved in unexpected heterotrophic processes includes those with a role in the degradation of polyaromatic hydrocarbons, use of polyhydroxyalkanoates as carbon-storage molecules and recycling of amino acids to produce compounds that can participate in redox reactions and energy production. Our study provides insights into how microorganisms in the plutonic crust are able to survive within fractures or porous substrates by coupling sources of energy to organic and inorganic carbon resources that are probably delivered through the circulation of subsurface fluids or seawater.

Diverse and abundant microorganisms are confirmed to be present in the basaltic upper crust^{5–7}, yet limited information exists about the distinctly different gabbroic lower crust that accounts for two-thirds of the volume of the crust^{8,9}. Expedition 360 of the International Ocean Discovery Program (IODP) drilled the Atlantis Bank oceanic core complex on the southwest Indian Ridge, Indian Ocean, where the lower oceanic crust is exhumed by detachment faulting to around 700 m below sea level, providing convenient access to an otherwise largely inaccessible realm (Fig. 1 and Supplementary Table 1).

Our approach combined cell counts, microscopy and shipboard quantification of ATP levels to assess the abundance and distribution of microbial communities, exoenzyme activity assays, carbon, hydrogen, nitrogen and sulfur (CHNS) and thin-section analyses of host rock samples, marker gene (iTAG) analyses of prokaryotic diversity, analyses of cultures to assess the viability and activities of selected taxa (including those that might be rare and may be missed by molecular approaches), and examination of expressed genes and lipid biomarkers. Rigorous efforts were made at all stages to control for contamination, including from drilling fluids (Supplementary Table 2) during shipboard sampling, and from laboratory and molecular kits (further details are provided in the Methods and Supplementary Information).

Samples along the 809-m depth of hole U1473A showed evidence of carbonate and/or clay-altered felsic veins (variable 0.01–6 mm

openings) within predominantly olivine gabbro, oxide gabbro and gabbro (Fig. 2) with around 0.1–5% porosity and 0.2–4 wt% water content. Borehole logging indicates temperatures of 15–18 °C. Total organic nitrogen (0.002–0.003 wt%) and total organic carbon (0.004–0.018 wt%) contents were extremely low (Supplementary Table 1).

Detection of biomass and biomarkers

Low-density (131–1,660 cells per cm³), heterogeneously distributed microbial communities were detected (Supplementary Table 3b), in the same range as recent studies of young, cool and oxic ridge-flank subsurface basalts in the Atlantic Ocean¹⁰ and samples from 0–15 m below seafloor (mbsf) obtained from the 3.5-million-year-old Atlantis Massif¹¹. Detection of archaeal intact polar lipids (IPLs) with a dominance of archaeol-based lipids over glycerol dialkyl glycerol tetraethers (GDGTs) indicates the existence of indigenous archaeal communities that are distinct from those that are observed in typical deep-biosphere sedimentary settings¹² (Fig. 3b, Extended Data Fig. 1 and Supplementary Table 3a). Bacterial diether glycerol (DEG) lipids were detected at all depths, which require less cellular maintenance than fatty-acid-containing lipids (Fig. 3c) and may indicate an adaptation to the stresses of living in the low-energy plutonic rocks of the deep biosphere¹³.

¹State Key Laboratory of Marine Geology, Tongji University, Shanghai, China. ²Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ³MARUM, University of Bremen, Bremen, Germany. ⁴Department of Geosciences, University of Bremen, Bremen, Germany. ⁵Department of Oceanography, Texas A&M University, College Station, TX, USA. ⁶Université de Brest, EA 3882, Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, ESIAB, Technopôle Brest-Iroise, Plouzané, France. ⁷Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ⁸Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA. ⁹These authors contributed equally: Jiangtao Li, Paraskevi Mara, Virginia P. Edgcomb. ✉e-mail: vedgcomb@whoi.edu

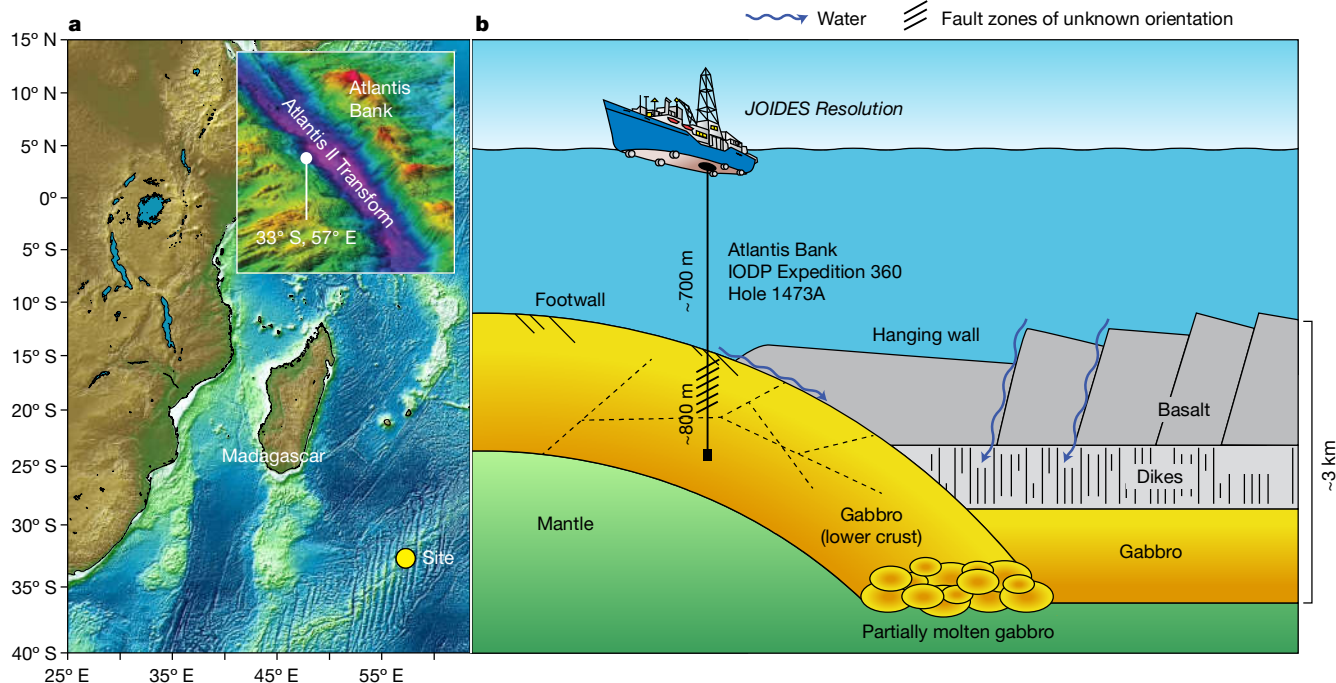


Fig. 1 | Study site of the exploration of the intrusive oceanic crust by IODP Expedition 360. a, Map of study site. **b**, Schematic of lower oceanic crust exhumed at the seafloor at Atlantis Bank. Dashed lines indicate the detected

presence of fractures along the depth of hole U1473A. The inset map is adapted from a previous study³¹; the background map was made using GeoMapApp (<http://www.geomapp.org/>).

The high IPL-to-core-lipid ratios support a likely in situ source of detected lipids. After cell death, the labile head groups of membrane lipids are enzymatically cleaved, leaving behind the core lipids that can be preserved in the rock record over millions of years¹⁴. Concentrations of core lipids were as low as for their IPL counterparts, ranging between 4 and 3,700 pg g⁻¹ of rock (Supplementary Table 3a). The resulting IPL-to-core-lipid ratios are high, on average more than 70% for glycosidic (G)-GDGTs/(G-GDGTs + core GDGTs) and 35% for glycosidic archaeols (G-ARs)/(G-ARs + core ARs), pointing to the minimal accumulation of fossil material. In comparison, IPLs typically comprise less than 10% of the sum of core lipids and IPLs in sedimentary settings¹².

Raman spectroscopy of a sample from 182 mbsf shows Raman bands that are consistent with amino acids and sugars. Scanning electron microscopy shows filamentous inclusions of organic compounds associated with iron and manganese oxides, surrounded by, and inter-grown with, calcite (CaCO₃), evidence of in situ origin (Fig. 3a). The viability of the cells is further supported by (1) the detection of methane production in long-term enrichments (average production from 9.62×10^{-4} – 5.66×10^{-2} nmol CH₄ g⁻¹ of rock per day) (Extended Data Fig. 2); (2) the cultivation of more than 100 unique fungal isolates from samples along the entire depth of the hole; (3) ATP concentrations above detection levels (1 pg cm⁻³) in 6 out of 11 rock samples (Fig. 2); and (4) detectable alkaline phosphatase exoenzyme activity rates of 0.04–2.3 pmol cm⁻³ h⁻¹ in all 5 samples for which measurements were attempted (Supplementary Table 3b).

Marker gene (iTAG) libraries reveal low diversity (Simpson scores (1–D) of 0.48–0.93) of putative heterotrophic and chemoautotrophic taxa, many of which have been described from deep-sea and polyextreme habitats (Extended Data Fig. 3). These include Nitrosopumilaceae, SAR202 Chloroflexi, Nitrospinaeae, SAR406 Marinimicrobia, SAR324 Deltaproteobacteria and Thioglobaceae, among other heterogeneously distributed taxa. Variation in the taxonomic composition among samples suggests that the availability of carbon and energy substrates is also heterogeneous. The SAR324 (marine group B) lineage of Deltaproteobacteria was detected in 8 out of 11 samples and is

known to have wide metabolic flexibility, including carbon fixation through Rubisco, and an ability to oxidize alkanes, methane and/or sulfur to generate energy^{15,16}—all potential metabolisms in the crustal basement. The detected DEG lipids could also be in part derived from these Deltaproteobacteria¹⁷. The putative sulfur and hydrogen oxidizers SAR406 Marinimicrobia were also detected in all samples¹⁸.

Non-metric multidimensional scaling (Extended Data Fig. 4) indicates that the measured ATP concentrations, fracturing intensity and graphic lithology all had the most-significant effect on observed depth variance in diversity for the 11 samples ($P \leq 0.05$). Locations of fracturing and alteration are where seawater from above, or low-temperature fluids from below, can most readily permeate the rock. Unfortunately, it was not possible to sample borehole fluids for important studies of in situ fluid chemistry during Expedition 360. Other studies of the subsurface lithosphere show that such fluids can carry the required substrates to support microbial life¹⁹.

mRNA reveals diverse strategies

Recovery of mRNA was lower for the shallowest samples from 11 and 168 mbsf, consistent with observations that the top of the drill hole was primarily dense unaltered gabbro with few fractures to support microbial populations (Supplementary Table 4; further details for all analyses are provided in the Supplementary Discussion). Detection of transcripts for various metabolic processes indicates that microorganisms have taken the prerequisite steps towards these processes. Additional measurements of activities, for example, through stable-isotope probing, in addition to those performed in this study are needed to demonstrate the occurrence of these processes. Expression of genes unique to peptidoglycan synthesis was detected in 5 out of 11 samples, indicating that cells are engaging in cell division and/or maintenance of cell integrity (Supplementary Table 4). Transcripts involved in biosynthesis of different amino acids were detected in 10 out of 11 samples as well as transcripts implicated in vitamin E and B12 synthesis (for example, *cbiC* and *cbiJ*) in 9 samples (Supplementary Table 4). The *cbiJ*

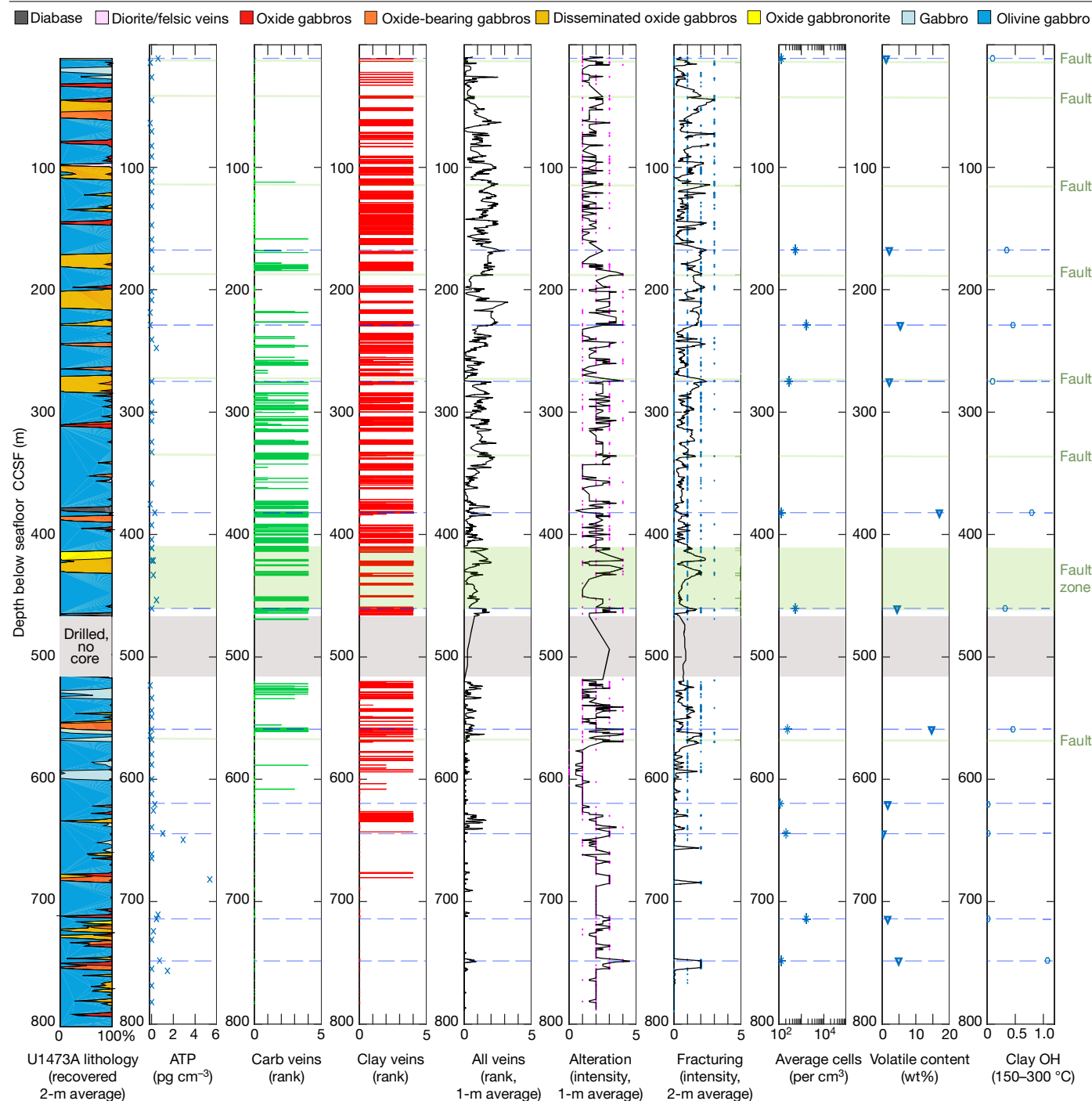


Fig. 2 | Summary of hole U1473A of IODP Expedition 360. Downhole properties of the gabbroic section determined by shipboard macroscopic descriptions and measured microbiological indicators. Labels indicate the property shown in each panel. Carb, carbonate. A key to lithology is included; rank and intensity range between 0 and 5. Faults are indicated by pale-green lines. Microbiological sample depths are highlighted by dashed blue lines. The

light-grey horizontal zone around 500 mbsf indicates a zone that was drilled but not cored. ATP was measured once per sample. CCSF, core composite depth below seafloor. Replicate cell-count data are provided in Supplementary Table 3b. Source data are available at [http://web.iodp.tamu.edu/DESCReport/\(file name, 360_U1473A_macroscopic.xlsx\)](http://web.iodp.tamu.edu/DESCReport/(file name, 360_U1473A_macroscopic.xlsx)).

sequence was annotated to *Chroococcidiopsis* sp., a cyanobacterial taxon that may contribute to anaerobic production of cobalamin or pseudocobalamin in the lower crust using a hydrogen-based lithoautotrophic metabolism²⁰.

Although we observed the production of methane in long-term enrichments (Extended Data Fig. 2), methanogens appear to be rare in our samples, explaining the low recovery of transcripts involved in methane metabolism (Supplementary Table 4), and no ribosomal

RNA gene sequences of known methanogens were recovered (Extended Data Fig. 3). However, the abundance of SAR324 Deltaproteobacteria in our iTAG libraries indicates potential methanotrophy^{15,16} coupled with methane production in our long-term enrichments (Extended Data Fig. 2). Some methane metabolism is probably present, but its importance is unclear in these samples. Detection of glycosidic archaeol and acyclic GDGT (GDGT-0) may indicate the presence of methanogens; however, these lipids may also derive from other archaeal sources.

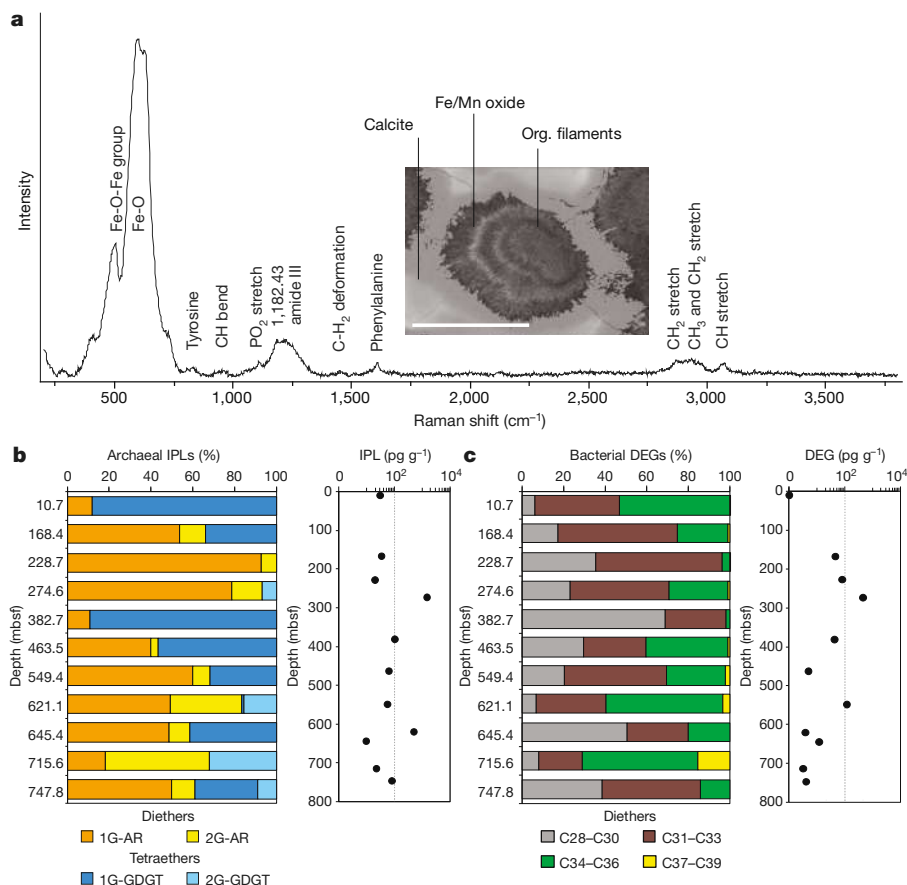


Fig. 3 | Biological signatures at hole U1473A revealed by Raman spectroscopy and membrane lipid analyses. **a**, Raman spectral features of organic inclusions in sample 21R2 from 182 mbsf. Inset, back-scattered electron image of an organic inclusion. Scale bar, 100 μm . Spectral measurements were performed at two spots on this particular feature, 10 independent times per spot, with similar results. **b**, Diversity (left) and concentration (right) of

archaeal IPLs within the 11 independent samples. **c**, Diversity (left) and concentration (right) of bacterial DEGs within the 11 independent samples. The summed carbon chain lengths from C28 to C39 are shown. 1G, monoglycosidic; 2G, diglycosidic; AR, archaeol. Lipid data are from single measurements owing to sample constraints. Source Data are available online.

Transcripts involved in nitrogen metabolism were detected in 9/11 samples, including genes associated with denitrification and dissimilatory nitrite reduction to ammonium, the latter of which was found in zones where fracturing and vein intensity were greatest (274.6, 460.4 and 619.6 mbsf). Transcripts involved in sulfur metabolism were detected in most samples (Supplementary Table 4).

Autotrophy might be expected in the gabbroic basement given the likelihood of only ephemeral sources of organic carbon via circulating fluids and the potential for seawater reactions with olivine-rich rocks that can produce molecular hydrogen (H_2), methane (CH_4) and other short-chain hydrocarbons that can be used as reducing agents for metabolic energy production¹⁹. Indeed, fluid inclusions present within gabbro from the Atlantis Bank represent a source of abiogenic CH_4 and H_2 , which can be released upon dissolution or fracturing of the mineral host²¹. Although the expression of genes that are potentially involved in 5 of the 6 known carbon-fixation pathways was observed (Supplementary Table 4), transcriptomes suggest that heterotrophy is also active, and that in situ microorganisms efficiently recycle and store the available organic compounds (Fig. 4). For example, the expression of transglycosylase genes associated with peptidoglycan degradation annotated to Desulfobacteriales and *Thermus* was detected at 460.4 and 643.9 mbsf. Furthermore, polycyclic aromatic hydrocarbons that can be formed through natural processes (for example, pyrolysis in hotter, deeper layers²²) may serve as carbon sources in the lower oceanic crust. We detected the expression of genes involved in polyaromatic and aromatic hydrocarbon degradation in 10 out of 11 samples; these

genes were primarily annotated to *Pseudomonas* (Supplementary Table 4). Recycled proteins are another important source of carbon. Expression of oligo-, DD-endo-, amino- and carboxypeptidases was detected (Supplementary Table 4). Expression of anabolic and catabolic genes implicated in the synthesis and degradation of amino acids known to be involved in Stickland reactions were detected in 10 of the 11 samples, accounting for up to 24% of normalized reads, indicating that these reactions may have a role in the nutrient-poor lower ocean crust (Fig. 4). Stickland reactions involve the oxidative deamination of one amino acid with a higher oxidation state, and the reductive deamination of another amino acid with a lower oxidation state, leading to ATP synthesis by substrate-level phosphorylation²³. Heterotrophic processes such as the degradation of amino acids may be the dominant terminal steps of organic carbon mineralization to CO_2 in these samples; however, the incomplete coverage of expressed genes in the recovered metatranscriptomes means that we cannot exclude contributions from additional pathways, including sulfate reduction, that may be expressed at low levels.

Many biochemical pathways produce acetyl-coenzyme A (acetyl-CoA), propionyl-CoA or acetoacetyl-CoA as intermediates that may contribute to the production of polyhydroxyalkanoates (PHAs)—bacterial and archaeal biopolymers known to serve as compatible solutes in halophiles²⁴. PHAs also serve as carbon-storage molecules, reducing equivalents and energy sources under carbon, nutrient and oxygen limitation^{25,26}. The primary role of PHAs in our samples is likely to be as energy storage in an environment in which the abundance of nutrients

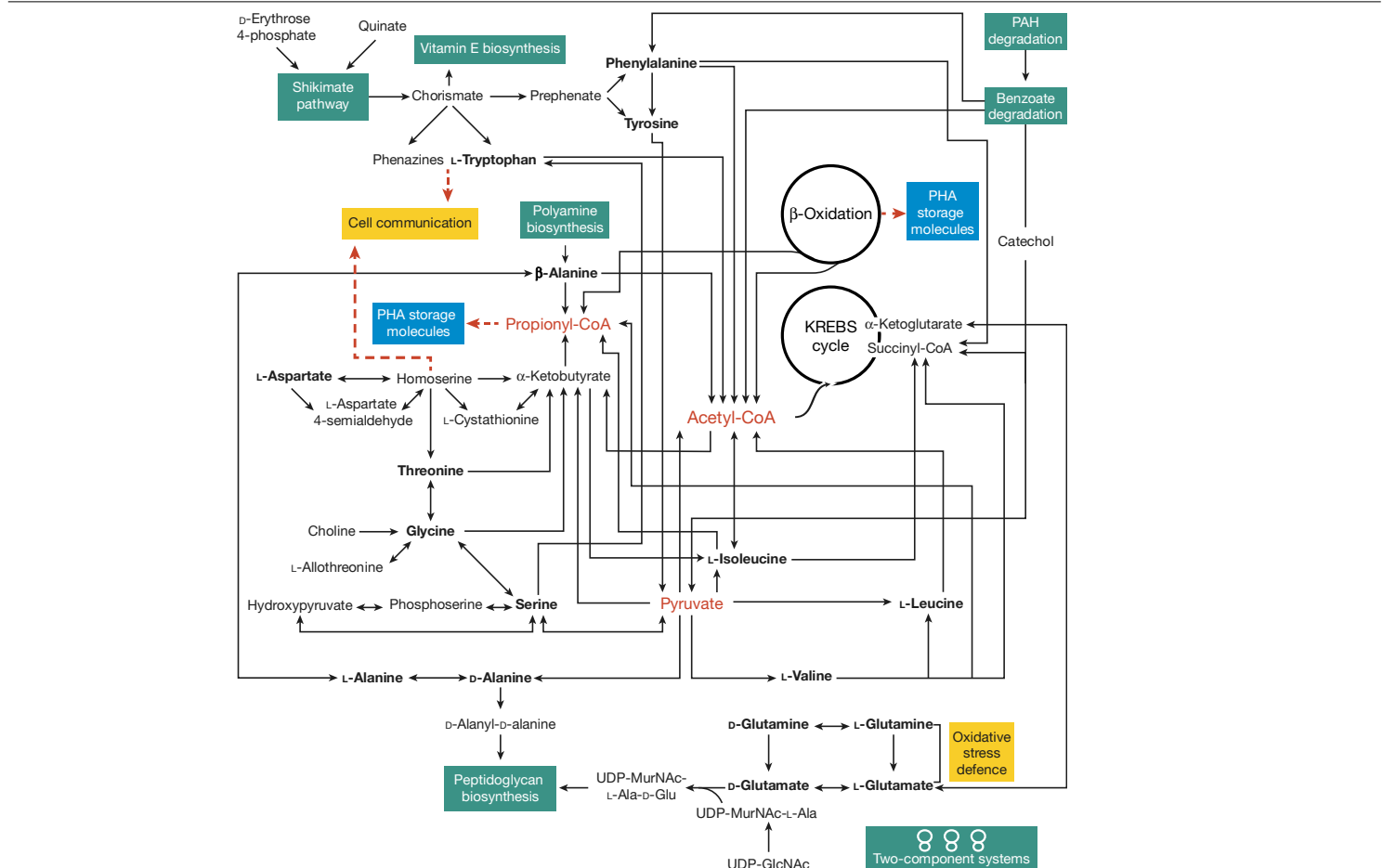


Fig. 4 | Schematic representation of metabolic processes inferred from observed transcripts in core samples from IODP Expedition 360. As the expression of all of the genes in the presented pathways was not detected in every sample, the schematic is presented as a working hypothesis only. PAH, polycyclic aromatic hydrocarbons; UDP-MurNAc, uridine diphosphate *N*-acetylmuramic acid; UDP-GlcNAc, uridine diphosphate *N*-acetylglucosamine. Amino acids are highlighted in bold. Blue boxes highlight PHA storage molecules, the production of which is supported by the detected transcripts.

Green boxes highlight biosynthetic processes, degradation pathways and two-component systems supported by transcripts. Yellow boxes highlight cellular activities supported by transcripts. Dashed red arrows represent expected (but unobserved) connections based on the data. The KREBS cycle (also known as the tricarboxylic acid cycle) generates energy via the oxidation of acetyl-CoA. β -Oxidation is the process of degradation of fatty acids to generate acetyl-CoA. The two processes have been circled.

can vary considerably over time. PHAs such as poly-3-hydroxybutyrate esters can be synthesized through an operon of three genes (*phaCAB*). Expression of *phaB* or *phaC* was detected in samples from 10.7 and 619.6 mbsf and both were detected in the sample from 714.9 mbsf; these genes were primarily annotated to *Pseudomonas*. Propionyl-CoA can be used for the biosynthesis of poly(3-hydroxybutyrate-co-3-hydroxyvalerate), a type of PHA that is produced by bacteria and archaea²⁷. Formation of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) requires the action of acetyl-CoA carboxylase and propionyl-CoA carboxylase. Expression of both genes was detected in the sample from 619.6 mbsf (Supplementary Table 4). β -Oxidation of fatty acids through 3-ketoacyl-CoA thiolase, 3-hydroxyacyl-CoA dehydrogenase and acyl-CoA dehydrogenase can produce propionyl-CoA, as can the catabolism of amino acids using 2-oxoisovalerate dehydrogenase and methylmalonyl-CoA mutase. One or more of these genes was detected in all samples, including those obtained from the two deepest samples (Fig. 4 and Supplementary Table 4). Collectively, our data suggest that PHAs have an important role in survival in the deep biosphere.

Inorganic phosphorus starvation is hypothesized to occur in the basaltic basement²⁸ and is a possible condition in the lower oceanic crust. The levels of inorganic phosphorus are efficiently regulated in bacteria by a two-component system²⁹ that involves ABC transporters.

Expression of phosphate transporters was detected in samples from 274.6, 460.4, 558.5, 619.6 and 747.7 mbsf. Alkaline phosphatase is part of the *pho* regulon that encodes extracellular enzymes capable of obtaining inorganic phosphorus. Genes of the *pho* regulon were expressed in samples from 619.6 and 643.9 mbsf, consistent with the detected alkaline phosphatase activity (Supplementary Table 3b). We hypothesize that recycled organic carbon may be used as a source of phosphorus for cells in this environment (Supplementary Tables 3, 4).

Our transcriptome data of life in the lower oceanic crust reveal heterotrophic activities that reflect the competition for limited and sporadically available resources, adaptations for withstanding long periods of resource scarcity, and efficient recycling of pools of organic matter in this challenging environment (further discussion of the results is provided in the Supplementary Discussion).

Conclusion

Circulation of fluids through fault zones and fractures in the lower oceanic crust can facilitate delivery of volatiles (for example, H₂, CO₂ and CH₄), nutrients, and abiotic and biotic electron donors and acceptors. Fluid pathways may represent advantageous habitats for microorganisms, yet the distribution of life in the deep crust remains

poorly understood. We reveal a low-biomass, but viable and/or active community in the exhumed lower oceanic crust, expanding our view of the extent of Earth's biosphere. Cellular activities appear to be very low based on enzyme activity measurements and mRNA recovery. Unexpectedly, heterotrophic processes may dominate over more-familiar autotrophic processes found at seafloor hydrothermal vents and in shallow marine sediments. Microorganisms can adapt to life in this 'slow lane' at least in part by using available fermentable organic molecules³⁰. Given the global expanse of the lower oceanic crust within known temperature limits for life, even low-biomass and slow-growing communities may make non-trivial contributions to global nutrient cycling. Future exploration of deeper lower ocean crust that is not exposed to faulting is required to determine whether the diversity and activities of microbiota present at those locations are similar to those found below the Atlantis Bank.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2075-5>.

- Shah Walter, S. R. et al. Microbial decomposition of marine dissolved organic matter in cool oceanic crust. *Nat. Geosci.* **11**, 334–339 (2018).
- D'Hondt, S., Rutherford, S. & Spivack, A. J. Metabolic activity of subsurface life in deep-sea sediments. *Science* **295**, 2067–2070 (2002).
- Jørgensen, B. B. Deep subseafloor microbial cells on physiological standby. *Proc. Natl Acad. Sci. USA* **108**, 18193–18194 (2011).
- Hoehler, T. M. & Jørgensen, B. B. Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* **11**, 83–94 (2013).
- Tully, B. J., Wheat, C. G., Glazer, B. T. & Huber, J. A. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *ISME J.* **12**, 1–16 (2018).
- Santelli, C. M., Edgcomb, V. P., Bach, W. & Edwards, K. J. The diversity and abundance of bacteria inhabiting seafloor lavas positively correlate with rock alteration. *Environ. Microbiol.* **11**, 86–98 (2009).
- Jungbluth, S. P., Bowers, R. M., Lin, H. T., Cowen, J. P. & Rappé, M. S. Novel microbial assemblages inhabiting crustal fluids within mid-ocean ridge flank subsurface basalt. *ISME J.* **10**, 2033–2047 (2016).
- Shrenk, M. O., Huber, J. A. & Edwards, K. J. Microbial provinces in the subseafloor. *Ann. Rev. Mar. Sci.* **2**, 279–304 (2010).
- Mason, O. U. et al. First investigation of the microbiology of the deepest layer of ocean crust. *PLoS ONE* **5**, e15399 (2010).
- Zhang, X., Feng, X. & Wang, F. Diversity and metabolic potentials of subsurface crustal microorganisms from the western flank of the Mid-Atlantic Ridge. *Front. Microbiol.* **7**, 363 (2016).
- Früh-Green, G. L. et al. Magmatism, serpentinization and life: insights through drilling the Atlantis Massif (IODP Expedition 357). *Lithos* **323**, 137–155 (2018).
- Lipp, J. S. & Hinrichs, K.-U. Structural diversity and fate of intact polar lipids in marine sediments. *Geochim. Cosmochim. Acta* **73**, 6816–6833 (2009).
- Valentine, D. L. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* **5**, 316–323 (2007).
- Summons, R. E. & Lincoln, S. A. in *Fundamentals of Geobiology* (eds Knoll, A. H.) 269–296 (John Wiley and Sons, 2012).
- Swan, B. K. et al. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**, 1296–1300 (2011).
- Sheik, C. S., Jain, S. & Dick, G. J. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ. Microbiol.* **16**, 304–317 (2014).
- Grossi, V. et al. Mono- and dialkyl glycerol ether lipids in anaerobic bacteria: biosynthetic insights from the mesophilic sulfate reducer *Desulfatibacillum alkenivorans* PF2803^T. *Appl. Environ. Microbiol.* **81**, 3157–3168 (2015).
- Hawley, A. K. et al. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. *Nat. Commun.* **8**, 1507 (2017).
- Kelley, D. S. et al. A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* **307**, 1428–1434 (2005).
- Puente-Sánchez, F. et al. Viable cyanobacteria in the deep continental subsurface. *Proc. Natl Acad. Sci. USA* **115**, 10702–10707 (2018).
- Klein, F., Grozeva, N. G. & Seewald, J. S. Abiotic methane synthesis and serpentinization in olivine-hosted fluid inclusions. *Proc. Natl Acad. Sci. USA* **116**, 17666–17672 (2019).
- Zolotov, M. & Shock, E. L. Abiotic synthesis of polycyclic aromatic hydrocarbons on Mars. *J. Geophys. Res. Planets* **104**, 14033–14049 (1999).
- Fonknechten, N. et al. *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomics* **11**, 555 (2010).
- Cai, L. et al. Comparative genomics study of polyhydroxyalkanoates (PHA) and ectoine relevant genes from *Halomonas* sp. TD01 revealed extensive horizontal gene transfer events and co-evolutionary relationships. *Microb. Cell Fact.* **10**, 88 (2011).
- Jendrossek, D. & Handrick, R. Microbial degradation of polyhydroxyalkanoates. *Annu. Rev. Microbiol.* **56**, 403–432 (2002).
- Liu, G. et al. Enoyl-CoA hydratase mediates polyhydroxyalkanoate mobilization in *Haloferax mediterranei*. *Sci. Rep.* **6**, 24015 (2016).
- Han, J. et al. Complete genome sequence of the metabolically versatile halophilic archaeon *Haloferax mediterranei*, a poly(3-hydroxybutyrate-co-3-hydroxyvalerate) producer. *J. Bacteriol.* **194**, 4463–4464 (2012).
- Lin, H.-T. et al. Inorganic chemistry, gas compositions and dissolved organic carbon in fluids from sedimented young basaltic crust on the Juan de Fuca Ridge flanks. *Geochim. Cosmochim. Acta* **85**, 213–227 (2012).
- Santos-Beneit, F. The Pho regulon: a huge regulatory network in bacteria. *Front. Microbiol.* **6**, 402 (2015).
- Zinke, L. A. et al. Thriving or surviving? Evaluating active microbial guilds in Baltic Sea sediment. *Environ. Microbiol. Rep.* **9**, 528–536 (2017).
- Dick, H. J. B. et al. The Atlantis Bank gabbro massif, Southwest Indian Ridge. *Prog. Earth Planet. Sci.* **6**, 64 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No power analyses were done to pre-estimate sample size. The sample size was determined by sample availability. The molecular control samples and actual samples were identified by different numbers and were treated blindly until the final stages during which these labels were replaced with the actual depth-associated identifiers.

Study site

The Atlantis Bank was uplifted from beneath sheeted dikes and pillow lavas in the local rift valley³² at the palaeo ridge-transform intersection. On the crest of the bank, there is a 25-km² wave-cut platform that is the location of an 809-m hole (U1473A) drilled during IODP Expedition 360 (32° 42.3402' S, 57° 16.6910' E) along a plate-spreading flow line at the centre of the gabbro massif. The eroded platform exposes massive foliated gabbro and oxide gabbro mylonites with enclaves of un-deformed olivine gabbro³¹. Oolitic sands and deep subaerial weathering occur on the flanks of the Bank down to 2,500 m depth, indicating the massif once rose to as much as 1.8-km above sea level, before subsiding to its present 700 m depth³¹, where it lies more than 3 km above the surrounding seafloor of the same age. Flexural uplift at the inside corner high can only account for around 1 km of this uplift, the rest is associated with later normal faulting accompanying isostatic uplift that occurred around 12 million years ago^{33,34}. A notable feature of the gabbro massif is that the boundary between crust and mantle rock is exposed along the transform valley wall for some 30 km. This structural geometry suggests that the overlying gabbro massif is in direct contact with partially serpentinized peridotite at a depth of around 4,500 m below site U1473A³³.

Sampling

Hole U1473A was drilled with a rotary coring bit, recovering cored rock in 10-m increments within core barrels that were brought on deck of the drill ship *JOIDES Resolution* and laid out on trays before splitting the cored rock samples into approximately 10–20-cm-long whole round core sections that were immediately selected and removed from the core liner for microbiological analysis by scientists who wore gloves and masks; each rock sample was placed into a sterile plastic bag for transit to the microbiology laboratory. Prioritized core samples exhibited evidence of low-temperature alterations by aqueous fluids, including carbonate veins and clay-altered felsic veins. We collected 68 samples for microbiological analysis; 11 of these—spanning the depth of hole U1473A—were selected for analyses from among those that showed the highest concentration of altered felsic and/or carbonate veins.

In the microbiology laboratory on-board the ship, each core section was rinsed three times with sterile distilled and deionized water, changing the bag each time, and then sprayed with 200-proof ethanol. At this point, after around 5 min, the core section was transferred to a custom laminar flow hood equipped with a HEPA filter and air supply that maintained positive air pressure. Within the hood, samples were photographed and the exterior approximately 1–2 cm of each whole round core was removed using a sterile chisel and rock hammer within a sterilized custom-made 0.3-m × 0.2-m × 0.1-m stainless-steel rock box. After removing core exteriors, core interiors were divided for different analyses. Three times during the cruise (during the first, second and third week of drilling), 1-litre samples of drilling fluid were filtered onto 45-mm 0.2-µm-pore-size Millipore Express Plus polycarbonate filters and frozen at –80 °C for DNA and RNA analysis. At the same time, 50-ml samples of drilling mud (Sepiolite) were also collected and frozen. Materials for DNA, RNA and lipid analyses were stored in sterile 50-ml Falcon tubes, whereas microscopy and thin-section samples were placed inside sterilized aluminium foil. These samples were carefully labelled, placed within plastic bags and immediately frozen at –80 °C. All equipment was rinsed and flame-sterilized between samples.

Contamination controls

Rotary coring contaminates the exteriors of core samples due to the circulation of drilling fluids (a mixture of Sepiolite and surface seawater) around the drilling bits. Extreme care was taken to remove or minimize this contamination and to not introduce new contaminations during sample handling and analysis. During IODP Expedition 360, a new less-volatile tracer, perfluoromethyldecalin, was successfully tested and calibrated and thus used to quantify the intrusion of drilling fluids into the interior of samples. As a further control for laboratory contamination, open Petri dishes containing microbiological medium used to culture fungi were placed inside the laminar flow hood during sample processing and these plates then were stored at room temperature for the duration of the cruise. In addition, extraction blanks, procedural controls and samples (of drilling fluids) were analysed for lipid and nucleic acid analyses (see the 'Lipid extraction and UHPLC-MS', 'DNA extraction and small subunit ribosomal-RNA marker-gene analysis' and 'RNA extraction and metatranscriptome analysis' sections).

Cell counts

Paraformaldehyde (4 ml; 4% solution in 100 mM phosphate-buffered saline) was added to autoclaved 7-ml plastic tubes. Then, 1 ml of powdered rock material was added to each tube, bringing the volume to 5 ml total. Two replicate tubes were prepared for cell counts for each sample and—where possible—vein material alone was aliquoted into one sample and whole-rock powder was aliquoted into the other. Preserved samples were stored at 4 °C for onshore analysis. For cell counts, 1 ml of the fixed sample slurry was used in the quantification procedure³⁵, which was modified to use 40 cycles of sonication instead of 20 to better release cells attached to the powdered rock. Cells were enumerated on filters stained with a 1:40 dilution of the stock SYBR Green I in TE buffer by counting either around 400 fields of view if fewer than a total of 40 cells were detected, or at least 40–50 cells in fewer fields when possible. The limit of quantification was defined as 3 × the s.d. of the mean of the negative-control counts. One negative control was processed and analysed for every 11 experimental samples. All samples were counted in duplicate.

ATP and exoenzyme assays

ATP concentration was assessed for all samples on board the *JOIDES Resolution* using luminescence methods³⁶ with the ATP Bioluminescence Assay Kit (Sigma-Aldrich) and a Turner Designs BioSystems 20/20 luminometer (Promega). The presence of ATP is indicative of microbial biomass. A standard curve of 0 (sterile MilliQ water), 1 and 100 ng l^{–1} ATP standard was run with each analysed set of samples. ATP in core samples was measured by placing 500 µl of ATP Assay Mix into a clean, sterile microcentrifuge tube after which the tube was incubated at room temperature for 3 min to allow hydrolysis of endogenous ATP, thus decreasing the background signal. Then around 0.5 cm³ of powdered sample was added (all tubes were weighed) to the same tube and the solution was immediately transferred to a clean 1.9-ml screw-top glass vial (acid washed) with the cap off, using a 1-ml pipette tip with the tip cut off. The vial was placed into the luminometer and results were read immediately. From the beginning of the expedition until drilling reached 218 mbsf, samples for ATP assays used leftover powdered rock in the steel rock box after separation of sample material for other assays. After observing largely undetectable ATP concentration, a second approach was used because microbial cells—if present—are likely to be more concentrated along cracks, within veins and vugs. Thus, starting at 218 mbsf, material from these specific features were included for these analyses.

Alkaline phosphatase activity was measured using fluorogenic substrate 4-methylumbelliferyl phosphate (Sigma-Aldrich) and its reference standard, methylumbelliferone. Fluorescence was measured using black, flat-bottom, 96-well microplates in a Spark 10M

Multimode Microplate Reader (Tecan). Fluorescence of methylumbelliferone is greatest at pH 10, therefore 25 μ l of 0.4 M NaOH was added to the wells (final concentration 40 mM) to be read. Then, 25 μ l of 1 M EDTA was added to the wells (100 mM final concentration) to prevent precipitation of carbonate from sampled veins³⁷. Fluorescence was measured with an excitation wavelength of 380 nm and emission of 454 nm for all substrates and standards. We mixed 1 cm³ powdered rock with 5 cm³ of sterile artificial seawater (ASW) in a 8-ml serum vial with 90:5:5 N₂:CO₂:H₂ headspace. Then, 700 μ l of each slurry was transferred with a sterile syringe into a 1.5-ml Eppendorf tube after set-up but before sealing the vial; this sample served as T_0 , which comprised triplicate 200 μ l technical replicates. The 700- μ l samples were briefly centrifuged (60 s at 2,500 rpm) and the supernatant was used for the T_0 analyses. Two additional samples were taken using the same methods as for T_0 after at least 2 weeks and then again after 4–6 weeks to generate a slope of activity. Incubations were kept at 10 °C, the inferred in situ temperature, for the duration of each assay.

Autoclaved, powdered rock from each of the samples was tested to determine the amount of fluorophore adsorbance to rock powder. Adsorbance was found to behave in a systematic way, resulting in a straight line when comparing fluorescence standards in ASW alone with fluorescence standards plus rock powder in ASW, although this relationship was found to be different when measured after 4 h versus days later. Therefore, a correction factor for adsorbance was applied to the enzyme data for the initial measurement (T_0 , $y = 1.90x - 676$), which was taken less than 2 h after experiment initiation, versus the second and third measurements (T_1 and T_2 , $y = 4.64x - 303$), which were taken days to weeks later. Negative controls consisting of the same ASW used for the sample incubations plus substrate, but no sample, were consistently below the limit of detection. The limit of quantification for the alkaline phosphatase assay, defined as 3 \times the s.d. of the blank, was 0.0242 pmol cm⁻³ of rock h⁻¹ based on the analysis of 8 blanks.

Carbon and nitrogen analysis

Powdered rock material from each sample (produced in the laminar flow hood on the *JOIDES Resolution* using a sterile mortar and pestle) was immediately transferred to sterile, muffled glass containers and stored in a desiccator until analysis of carbon and nitrogen according to established methods^{38,39}. In brief, samples were weighed into methanol-rinsed silver boats (4 mm \times 6 mm, Costech). Then, 96-well glass plates (combusted for 4 h at 450 °C) holding these samples were placed in a vacuum desiccator that also contained an open dish with about 50 ml fresh, concentrated (12 N) HCl. An inverted crystallization dish was placed over the samples to protect them from water condensation. The desiccator was closed and pumped out with an air-driven aspirator, to a reading of about 0.5 atm and the desiccator was placed in an oven kept between 60 and 65 °C. Acidification was allowed to run for 60 to 72 h, as described previously⁴⁰. When acidification was complete, the samples were removed and set in the oven to dry (60–65 °C). Subsequently, the samples were placed in a vacuum desiccator charged with indicating silica gel (Fisher S162-500, activated by heating above 220 °C for several hours) and pumped down again and dried for about 24 h. Samples were then analysed on a Costech 4010 Elemental Analyzer connected via a Finningan-MAT ConFlo-II interface to a DeltaPlus isotope ratio mass spectrometer.

Thin-section preparation, scanning electron microscopy and Raman analyses

Thin-section billets were cut from dedicated subsamples of the 11 samples examined in this study. Thin sections were prepared by High Mesa Petrographics. Mosaic images were taken of all thin sections in transmitted and reflected light and were used to guide the scanning electron microscopy (SEM) and Raman analyses. Uncoated thin sections were screened with SEM in low-vacuum mode to search for and visualize carbon-rich inclusions of possible organic origin. SEM was performed

on a Hitachi TM3000 scanning electron microscope equipped with a Bruker energy-dispersive spectroscopy system for imaging and semi-quantitative element analysis. Promising samples with possible organic inclusions were analysed using a computer-controlled, high-resolution confocal Raman system (Horiba LabRam HR) equipped with three lasers (633 nm, 532 nm and 473 nm), a motorized stage and a SWIFT fast-mapping option. Confocal Raman spectroscopy enables the non-destructive analysis and recognition of living and fossil (once-living) microorganisms in altered igneous rocks. The achievable lateral and spectral resolution of this instrument is better than 1 μ m and 2 cm⁻¹, respectively. Spectra were analysed using the BioRad KnowItall software and spectral databases to identify organic compounds.

Lipid extraction and UHPLC–MS lipid biomarker analysis

Crushed core samples stored in Falcon tubes at –80 °C were first milled for 10 min to a fine powder and subsequently extracted with a modified Bligh and Dyer method according to a previous study⁴¹. Before milling and extraction of each sample, a procedure blank was performed. First, a milling blank was performed using combusted sea sand (fired at 450 °C for 5 h) to clean the mill and to limit cross-contamination of samples. Subsequently, this sea sand was transferred to geo-cleaned (rinsed three times with a mixture of methanol and dichloromethane) Teflon containers used for the extraction of the samples and solvent-extracted in the same manner as the samples. For this, 100 ng of an internal standard (C46 glycerol trialkyl glycerol tetraether) and around 50 ml of a solvent mixture of dichloromethane:methanol:buffer (2:1:0.8, v/v) was added to the sample in the Teflon container and ultrasonicated for 10 min using a geo-cleaned ultrasonic stick. After ultrasonication, the samples were centrifuged (1,750 rpm at 10 min) and the supernatant was transferred to a fired separatory funnel. The samples were extracted in four steps, for the first two steps a phosphate buffer (K₂HPO₄, 50 mM at pH 7.4) was used, in the third step the phosphate buffer was replaced by 5% trichloroacetic acid (50 g l⁻¹ at pH 2) and in the last step only dichloromethane:methanol (9:1, v/v) was used. Equal amounts of dichloromethane and deionized MilliQ water were added to the extract collected in the separatory funnel, the mixture was shaken, and the organic phase was collected as the total lipid extract and blown to dryness under a gentle stream of nitrogen.

An aliquot of the total lipid extract was analysed using ultrahigh-pressure liquid chromatography (UHPLC) coupled to mass spectrometry (MS) on a Dionex Ultimate 3000RS UHPLC connected to an ABSciEX QTRAP4500 Triple Quadrupole/Ion Trap MS (UHPLC-triple quad-MS) using a Turbolon electrospray ion (ESI) source. Separation of compounds was achieved on a Waters Acquity BEH C18 column (1.7 μ m, 2.1 mm \times 150 mm) equipped with a guard column of the same material following a previously published protocol⁴². Compounds of interest were screened with multiple-reaction monitoring and selected-ion monitoring techniques as described previously⁴². Concentrations of lipids were determined relative to the internal C46 glycerol trialkyl glycerol tetraether standard and were corrected for individual response factors using commercially available standards (diC16-DEG, archaeol) and isolated standards from cultures (GDGT-0, 1G-AR, 2G-AR, 1G-GDGT-0 and 2G-GDGT-0). Sciex Analyst 1.6.3 and Sciex MultiQuant 3.0.3 (AB Sciex) were used for triple-quadrupole MS data acquisition and data processing.

The presence of crenarchaeol was confirmed by core GDGT analysis according to a previously published study⁴³. In brief, an aliquot of the total lipid extract was analysed using a Dionex Ultimate 3000RS UHPLC connected to a Bruker maXis ultrahigh-resolution quadrupole time-of-flight MS, equipped with an atmospheric pressure chemical ionization (APCI) II source. Compounds were separated using two acquity BEH HILIC amide columns (1.7 μ m, 2.1 mm \times 300 mm) in tandem maintained at 50 °C, and *n*-hexane as eluent A and *n*-hexane:isopropanol (90:10, v/v) as eluent B (a detailed protocol has been published previously⁴³). Drilling mud and extraction blank contamination controls were also

run for lipid biomarker analyses (Extended Data Figs. 5, 6). Bruker Compass 1.9 and Bruker data analysis v.4.4 (Bruker Daltonics) were used for quadrupole time-of-flight data acquisition and processing.

DNA extraction and small subunit ribosomal-RNA marker-gene analysis

Rock material was crushed while still frozen in a Progressive Exploration Jaw Crusher (Model 150), the surfaces of which were sterilized with 70% ethanol and RNase AWAY (Thermo Fisher Scientific) inside a laminar flow hood. Powdered rock material was returned to the -80°C freezer until extraction. DNA was extracted from 20, 30 or 40 g of powdered rock material, depending on the quantity of rock available. A DNeasy PowerMax Soil Kit (Qiagen) was used according to the manufacturer's protocol modified to include three freeze–thaw treatments before the addition of Soil Kit solution C1. Each treatment consisted of 1 min in liquid nitrogen followed by 5 min at 65°C . DNA extracts were concentrated by isopropanol precipitation overnight at 4°C . The low biomass in our samples required whole-genome amplification before PCR amplification of marker genes. Genomic DNA was amplified by multiple displacement amplification using the REPLI-g Single Cell Kit (Qiagen) as described. Multiple displacement amplification bias was minimized by splitting each whole-genome amplification sample into triplicate 16 μl reactions after 1 h of amplification and then resuming amplification for the manufacturer-specified 7 h (8 h total). DNA was also recovered from samples of drilling mud and drilling fluid (surface water collected during the coring process) for negative controls, as well as two 'kit control' samples, in which no sample was added, to account for any contaminants originating from either the DNeasy PowerMax Soil Kit or the REPLI-g Single Cell Kit. Bacterial small subunit ribosomal RNA (rRNA) gene fragments were PCR amplified from multiple displacement amplification samples and sequenced at the Georgia Genomics and Bioinformatics Core (University of Georgia). The primers used were: Bac515-Y and Bac926R⁴⁴. Dual-indexed libraries were prepared with (HT) iTruS (Kappa Biosystems) chemistry and sequencing was performed on an Illumina MiSeq 2 \times 300-bp system with all samples combined equally on a single flow cell.

Raw sequence reads were processed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), FLASH (<http://ccb.jhu.edu/software/FLASH/>) and FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) for trimming and removal of low-quality and/or short reads. Quality filtering included requiring a minimum average quality of 25 and rejection of paired reads that were less than 250 nucleotides. Operational taxonomic unit (OTU) clusters were constructed at 99% similarity with the script `pick_otus.py` within the Quantitative Insights Into Microbial Ecology (QIIME) v.1.9.1 software⁴⁵ and `ucrust`. Any OTU that matched an OTU in one of our control samples (drilling fluids, drilling mud, extraction and whole-genome amplification controls) was removed (using `filter_otus_from_otu_table.py`) along with any sequences of land plants and human pathogens that may have survived the control filtering due to clustering at 99% (`filter_taxa_from_otu_table.py`). As an additional quality-control measure, genera that are commonly identified as PCR contaminants were removed^{46,47}.

Unclassified OTUs were queried using BLAST against the GenBank nr database and further information about these OTUs is provided in the Supplementary Discussion. OTUs that could not be assigned to Bacteria or Archaea were removed from further analysis. For downstream analyses, any OTUs not representing more than 0.01% of the relative abundance of sequences overall were removed as those are unlikely to contribute considerably to in situ communities. The OTU data table was transformed to a presence or absence table and the Jaccard method was used to generate a distance matrix using the `dist.binary()` function in the R package `ade4`. A hierarchical clustering dendrogram was created using `hclust()` and the stability of the clusters was evaluated using the `clusterboot()` function in the `fpc` package in R with 500 iterations.

RNA extraction and metatranscriptome analysis

Frozen rock material was crushed as above and then ground quickly into a fine powder using a precooled sterilized mortar and pestle, after which RNA extraction started immediately. The jaw crusher was cleaned and rinsed with 70% ethanol and RNaseZap RNase Decontamination Solution (Invitrogen) between samples. About 40 g of material was extracted for each sample using the RNeasy PowerSoil Total RNA Isolation Kit (Qiagen) according to the manufacturer's protocol with the following modifications. Each sample was evenly divided into 8 bead tubes (Qiagen) and then 2.5 ml of Bead solution was added into the bead tube followed by 0.25 ml of solution SR1 and 0.8 ml of solution SR2. Bead tubes were frozen in liquid nitrogen and then thawed at 65°C in a water bath three times. RNA was purified using the MEGAclean Transcription Clean-up Kit (Ambion) and concentrated with an overnight isopropanol precipitation at 4°C . Trace amounts of contaminating DNA were removed from the RNA extracts using TURBO DNA-free (Invitrogen, USA) as described by the manufacturer. To ensure that DNA was removed thoroughly, each RNA extract was treated twice with TURBO DNase (Invitrogen). A nested PCR reaction (2 \times 35 cycles) using bacterial primers⁴⁸ was used to confirm the absence of DNA in our RNA solutions.

RNA was converted to cDNA using the Ovation RNA-Seq System V2 kit (NuGEN) according to the manufacturer's protocol to preferentially prime non-rRNA sequences. The cDNA was purified with the MinElute Reaction Cleanup Kit (Qiagen) and eluted into 20 μl elution buffer. Extracts were quantified using a Qubit Fluorometer (Life Technologies) and cDNAs were stored at -80°C until sequencing using 150-bp paired-end reads on an Illumina NextSeq 550.

To control for potential contaminants introduced during drilling, sample handling and laboratory kit reagents, we sequenced a number of control samples as described above. This included two samples that controlled for potential nucleic acid contamination, a 'method' control to monitor possible contamination from our laboratory extractions, which included around 40 g sterilized glass beads processed through the entire protocol in place of rock, and a 'kit' control to account for any signal coming from trace contaminants in kit reagents, which received no addition. In addition, three field controls were extracted: a sample of the drilling mud (Sepiolite), and two drilling seawater samples collected during the first and third weeks of drilling. cDNA obtained from these controls was sequenced together with the rock samples and co-assembled.

Trimmomatic (v.0.32)⁴⁹ was used to trim adaptor sequences (leading = 20, trailing = 20, sliding window = 04:24, minlen = 50). Paired reads were further quality checked and trimmed using FastQC (v.0.11.7) and FASTX toolkit (v.0.014). Downstream analyses used paired reads. After co-assembling reads with Trinity (v.2.4.0) from all controls (minimum length, 150 bp), Bowtie2 (v.2.3.4.1)⁵⁰ was used (with the parameter 'un-conc') to align all sample reads to this co-assembly. Reads that mapped to our control co-assembly allowing 1 mismatch were removed from further analysis (23.5–68.5% of sequences remained in sample datasets; Supplementary Table 4). Trinity (v.2.4.0) was used for de novo assembly of the remaining reads in sample datasets (minimum length, 150 bp). Bowtie aligner was used to align reads to assembled contigs, RSEM⁵¹ was used to estimate the expression level of these reads, and TMM was used to perform cross-sample normalization and to generate a TMM-normalized expression matrix. Within the Trinotate suite, TransDecoder (v.3.0.1) was used to identify coding regions within contigs and functional and taxonomic annotation was made by BLASTx and BLASTp against UniProt, Swissprot (release 2018_02) and RefSeq non-redundant (nr) protein sequence databases (e-value threshold of 1×10^{-5}). BLASTp was used to look for sequence homologies with the same e values. HMMER (v.3.1b2) was used to identify conserved domains by searching against the Pfam (v 31.0) database⁵². SignalP (v.4.1)⁵³ and TMHMM (2.0c)⁵⁴ were used to predict signal peptides and transmembrane domains. RNAMEMER (v.1.2)⁵⁵ was used to identify

rRNA homologies of archaea, bacteria and eukaryotes. Because the Swissprot database does not have extensive representation of protein sequences from environmental samples, particularly deep-sea and deep-biosphere samples, annotations of contigs used for analyses of selected processes that we report were manually cross checked by BLASTx against the GenBank nr database.

Aside from removing any reads that mapped well to our control co-assembly (1 mismatch), as an extra precaution, any sequence that exhibited $\geq 95\%$ sequence identity over $\geq 80\%$ of the sequence length to suspected contaminants (for example, human pathogens, plants or taxa known to be common molecular kit reagent contaminants, and not described from the marine environment) as described previously^{46,56} were removed. This conservative approach potentially removed environmentally relevant data that were annotated to suspected contaminants due to poor taxonomic representation from environmental taxa in public databases; however, it affords the highest possible confidence about any transcripts discussed. Additional functional annotations of contigs were obtained by BLAST against the KEGG, COG, SEED and MetaCyc databases using MetaPathways (v.2.0) to gain insights into particular cellular processes, and to provide overviews of metabolic functions across samples based on comparisons of fragments per kilobase per million (FPKM)-normalized data. All annotations were integrated into a SQLite database for further analysis.

Statistics

Ocean-drilling legs with diverse sampling requirements can provide only a limited volume of rock material for microbiology from each depth horizon. Although still providing tremendously valuable insights into the deep lithified biosphere, the absence of replicate samples limits the types of statistical analyses that we can perform. A clustering analysis of curated transcripts within the functional categories presented in this paper was performed using log + 1-transformed FPKM values and the Ward method, and the distance matrix was constructed using the Manhattan method and pvcust in R. Non-metric multidimensional scaling (NMDS) analysis was performed on the Jaccard distance matrix of prokaryotic OTU presence or absence data for 11 samples but not on transcript data because we report here only a carefully curated subset of total reads for each sample.

Taxonomic assignments of discussed contigs are presented in Supplementary Table 4, and were selected from among the top-10 BLASTx hits. In cases where different taxa were included within the top-10 hits for a conserved domain, either the top hit or the top consensus hit to a known marine group was selected.

Fungal culturing

Culturing efforts to specifically isolate fungi were performed using several culture conditions using not only classical culture-based approaches but also laser-nephelometry-based high-throughput culturing. Three different culture media were used (a modified Sabouraud (1 g l⁻¹ mycological peptone, 3% sea salts, 15 g l⁻¹ agar), a yeast nitrogen base (7 g l⁻¹ YNB powder (Difco), 3% sea salts, 15 g l⁻¹ agar), a minimal medium as defined previously⁵⁷ and formulated with or without crushed oceanic crustal rock (10 g l⁻¹), with variable carbon sources (galactose, celluloses or chitin at 15 g l⁻¹), with a mix of essential amino acids, and with or without a mix of antibiotics (chloramphenicol and penicillin G). Each experiment was processed at different temperatures ranging from 5 to 35 °C. After isolation, a dereplication step by mini- or microsatellite-primed (MSP)-PCR was processed to obtain unique isolates⁵⁸ that were identified based on sequencing of several genetic markers (ITS, 18S, 28S or 26S rRNA genes, RPB2, tubulin and actin genes).

Methane analysis

Roughly 5 cm³ of rock from depths of 247.71, 279.55, 621.09, 724.68 and 747.78 mbsf were added to sterile ASW to 27 ml total in 30-ml

serum vials. The control for the methane measurements was a vial of 27 ml of ASW with no rocks added. The vials were sealed with butyl stoppers and gassed with a mixture of 90% N₂, 5% H₂ and 5% CO₂, and incubated at 10 °C. Methane in the headspace of the vials was measured at 25 weeks and 60 weeks after incubation. The gas in the headspace of the vials was extracted and measured using gas chromatography–flame ionization detection (GC–FID), following protocols similar to previously described methods⁵⁹. Calibrations were performed for the measurements using methane tanks at 0 ppm, 1.0 ppm and 6.41 ppm, and the GC–FID loop was flushed for 1 min before each calibration measurement. The syringe used to inject the gas samples into the GC–FID loop was flushed three times with the 90% N₂, 5% H₂ and 5% CO₂ gas mixture before each sample extraction and injection.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

iTAG data for the 11 samples as well as 13 different negative controls (drilling muds and fluids, seawater and kit controls) are deposited in the NCBI BioProject under accession number PRJNA497074. Results are presented at approximately phylum level; however, taxonomic assignments at finer resolution are available from the corresponding author upon request. Raw reads for transcriptome data have also been deposited in BioProject under accession number PRJNA497074; iTAGs have been deposited in the NCBI SRA under accession numbers SRR8136794–SRR8136814 and transcript raw reads can be found in the SRA under accession numbers SRR8141073–SRR8141077. Assemblies for curated portions of the data presented are available upon request to the corresponding author. All relevant data are available from the corresponding author or are included with the manuscript as Supplementary Information, and the Source Data for Fig. 2 are available at <http://web.iodp.tamu.edu/DESCReport/> (file name, 360_U1473A_macroscopic.xlsx). Source Data for Fig. 3 and Extended Data Figs. 1, 2, 6 are provided with the paper.

32. Fox, P. J. & Gallo, D. G. A tectonic model for ridge-transform-ridge plate boundaries: implications for the structure of oceanic lithosphere. *Tectonophysics* **104**, 205–242 (1984).
33. Dick, H. J. B. et al. Dynamic accretion beneath a slow-spreading ridge segment: IODP hole 1473A and the Atlantis Bank oceanic core complex. *J. Geophys. Res. Solid Earth* **124**, 12631–12659 (2019).
34. Baines, A. G. et al. Mechanism for generating the anomalous uplift of oceanic core complexes: Atlantis Bank, southwest Indian Ridge. *Geology* **31**, 1105–1108 (2003).
35. Morono, Y., Terada, T., Kallmeyer, J. & Inagaki, F. An improved cell separation technique for marine subsurface sediments: applications for high-throughput analysis using flow cytometry and cell sorting. *Environ. Microbiol.* **15**, 2841–2849 (2013).
36. Lundin, A., Hasenson, M., Persson, J. & Pousette, A. Estimation of biomass in growing cell lines by adenosine triphosphate assay. *Methods Enzymol.* **133**, 27–42 (1986).
37. Coolen, M. J. & Overmann, J. Functional exoenzymes as indicators of metabolically active bacteria in 124,000-year-old sapropel layers of the eastern Mediterranean Sea. *Appl. Environ. Microbiol.* **66**, 2589–2598 (2000).
38. Pella, E. Elemental organic analysis. Part 1, historical developments. *Am. Lab.* **22**, 116–125 (1990).
39. Pella, E. Elemental organic analysis. Part 2: State of the art. *Am. Lab.* **22**, 28–32 (1990).
40. Whiteside, J. H. et al. Pangean great lake paleoecology on the cusp of the end-Triassic extinction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **301**, 1–17 (2011).
41. Sturt, H. F., Summons, R. E., Smith, K., Elvert, M. & Hinrichs, K. U. Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry—new biomarkers for biogeochemistry and microbial ecology. *Rapid Commun. Mass Spectrom.* **18**, 617–628 (2004).
42. Klein, A. T. et al. Investigation of the chemical interface in the soybean–aphid and rice–bacteria interactions using MALDI-mass spectrometry imaging. *Anal. Chem.* **87**, 5294–5301 (2015).
43. Becker, K. W. et al. An improved method for the analysis of archaeal and bacterial ether core lipids. *Org. Geochem.* **61**, 34–44 (2013).
44. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).

45. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
46. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
47. Sheik, C. S. et al. Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the Census of Deep Life. *Front. Microbiol.* **9**, 840 (2018).
48. Cole, J. R. et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**, D141–D145 (2009).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
51. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
52. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
53. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
54. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
55. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
56. Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. & Chiodini, R. J. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* **8**, 24 (2016).
57. Le Calvez, T., Burgaud, G., Mahé, S., Barbier, G. & Vandenkoornhuyse, P. Fungal diversity in deep-sea hydrothermal ecosystems. *Appl. Environ. Microbiol.* **75**, 6415–6421 (2009).
58. Burgaud, G., Arzur, D., Durand, L., Cambon-Bonavita, M.-A. & Barbier, G. Marine culturable yeasts in deep-sea hydrothermal vents: species richness and association with fauna. *FEMS Microbiol. Ecol.* **73**, 121–133 (2010).
59. Valentine, D. L. et al. Propane respiration jump-starts microbial response to a deep oil spill. *Science* **330**, 208–211 (2010).

Acknowledgements We thank the captain, crew and all who sailed on *JOIDES Resolution* for IODP Expedition 360, whose support was essential; T. Sehein and M. Torres-Beltran for assistance with statistical analyses; Q. Ma, E. S. Taylor (WHOI Creative) and Andrew Newman Design for assistance with figures; S. Yvon-Lewis for providing materials and assistance with methane measurements; and M. Sogin, J. Huber, B. Orcutt, K. Lloyd, J. Biddle and S. D'Hondt for helpful discussions on the relative merits of different molecular data-handling options for contamination controls. This study was funded by National Science Foundation grants OCE-1658031 to V.P.E. and F.K., OCE-1658118 to J.B.S., and OCE-1450528 and OCE-1637130 to H.J.B.D. F.S. acknowledges funding from the DFG under Germany's Excellence Strategy (no. EXC-2077-390741603) and the Gottfried Wilhelm Leibniz Program (HI 616-14-1). Support to J.L. was provided by the National Science Foundation of China (no. 41772358) and the Ministry of Science and Technology of China (no. 2012CB417302).

Author contributions V.P.E., J.B.S., F.K., F.S. and J.L. acquired funding. V.P.E. and J.B.S. collected samples, performed shipboard assays and established enrichment cultures. J.L. extracted RNA. P.M., J.L. and V.P.E. performed mRNA analyses. D.B., S.L. and R.C. extracted DNA, and D.B. and S.Y.W. analysed iTAG data. F.S. and L.A.E.M. performed lipid biomarker analyses. G.B. and M.Q. performed fungal isolations. F.K. performed Raman spectroscopy experiments and J.L. carried out SEM with F.K. J.B.S. and S.Y.W. performed cell counts and exoenzyme assays, and analysed methane-generation experiments. H.J.B.D. and F.K. provided the geological context, and D.K.B. compiled downhole-core description data. P.M. conceived the working hypotheses in Fig. 4. V.P.E. wrote the manuscript draft and all authors contributed to review and editing.

Competing interests The authors declare no competing interests.

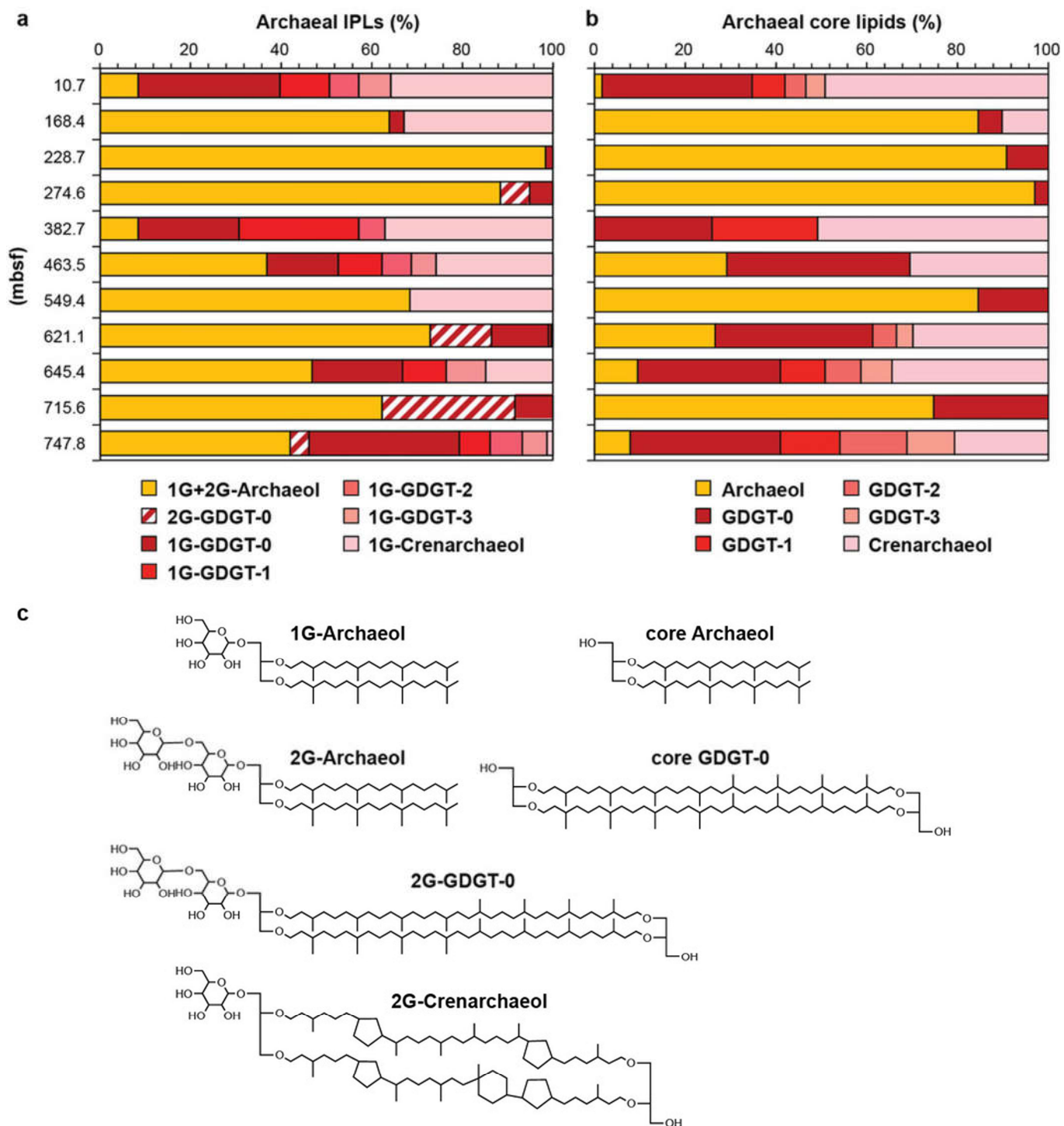
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2075-5>.

Correspondence and requests for materials should be addressed to V.P.E.

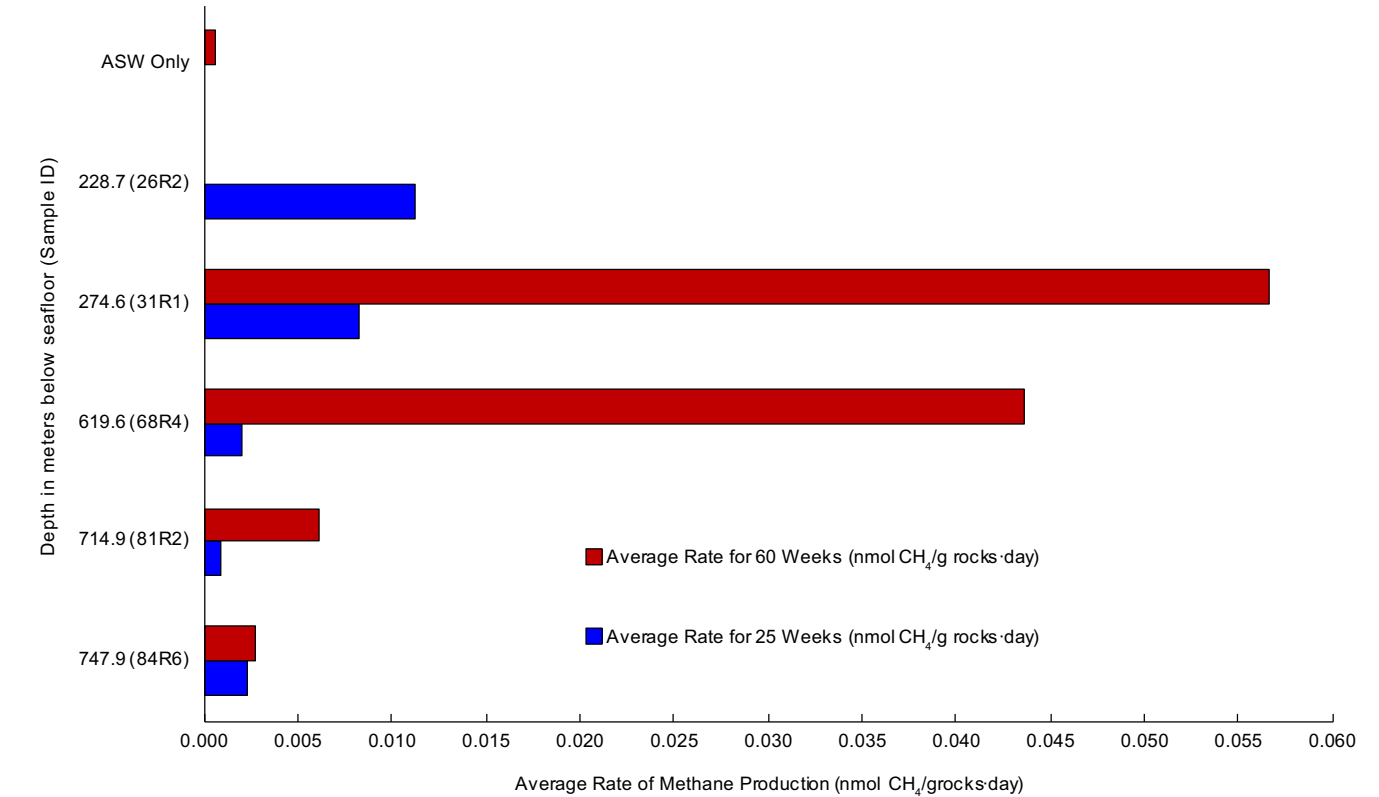
Peer review information *Nature* thanks Bo Barker Jørgensen, Jennifer Biddle and Steven D'Hondt for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



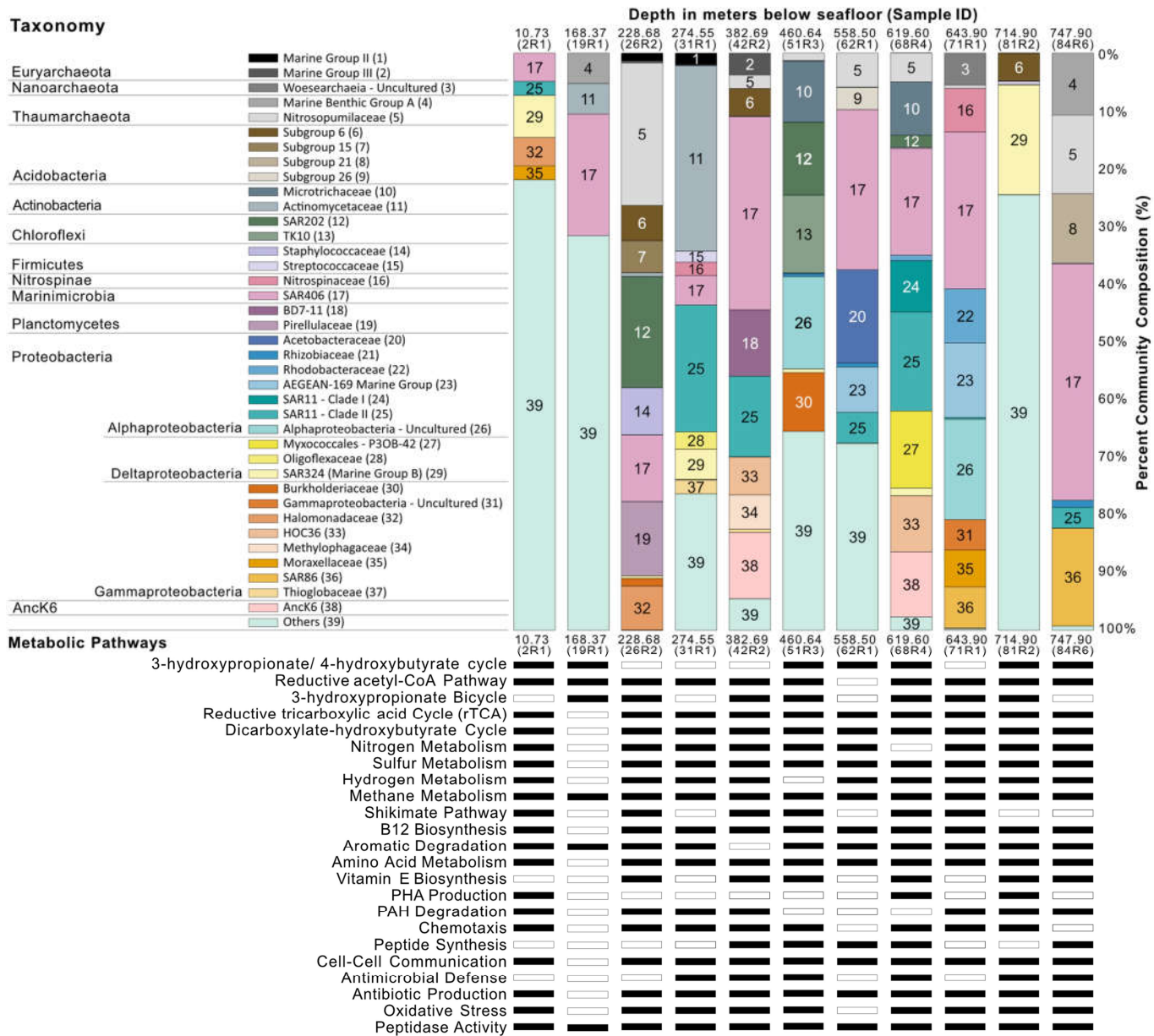
Extended Data Fig. 1 | Archaeal core lipid analyses. a, b, Downhole changes in core lipid composition of archaeal IPLs (**a**) and archaeal core lipids (**b**). The relative abundances of the diether lipid archaeol and tetraether lipids with zero, one, two and three rings (GDGT-0, GDGT-1, GDGT-2 and GDGT-3,

respectively) and crenarchaeol are shown. **c,** Structures of the most abundant archaeal lipids. 1G, monoglycosidic; 2G, diglycosidic; AR, archaeol; GDGT, glycerol dialkyl glycerol tetraether. Lipid data are from single measurements owing to sample constraints. Source Data are available online.



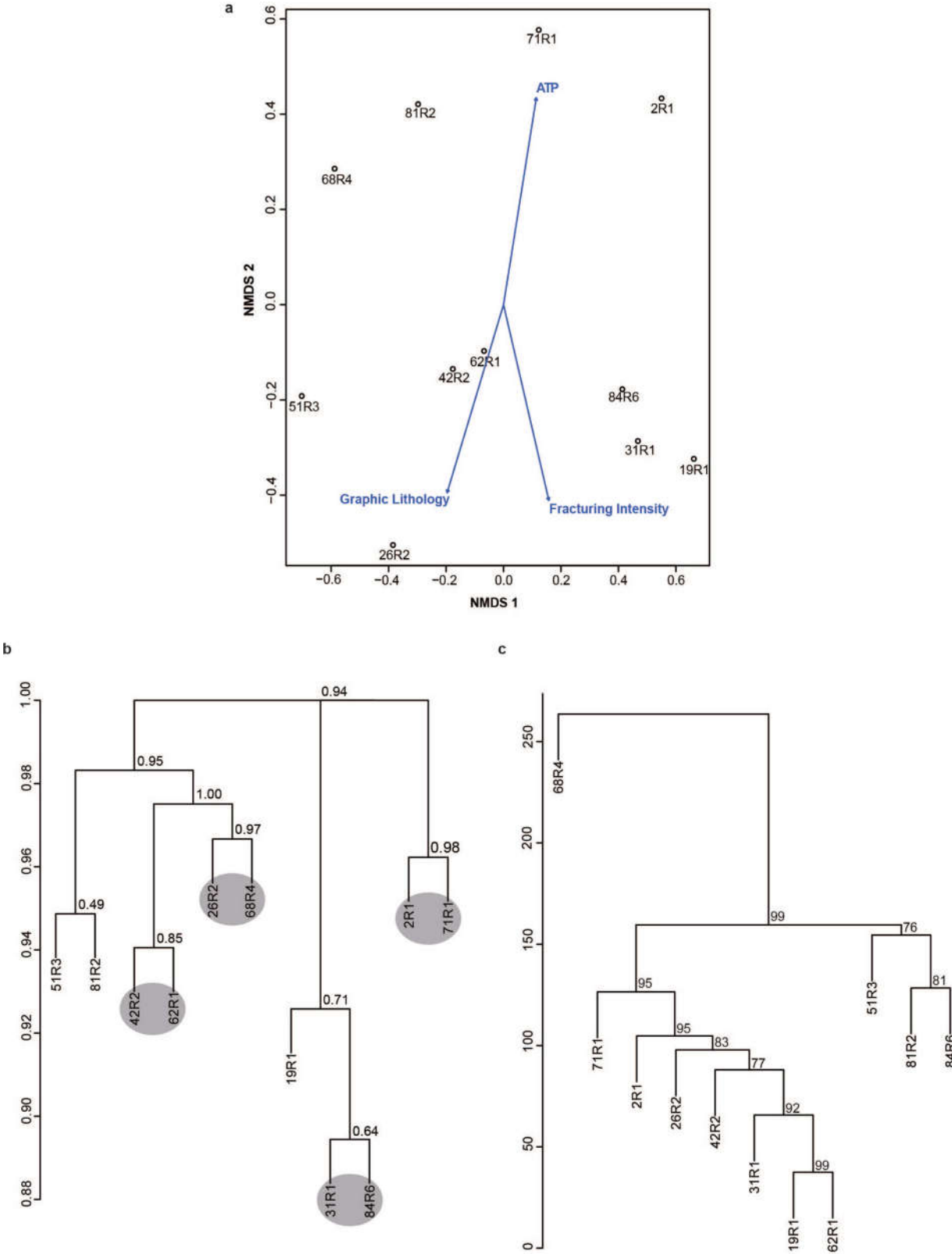
Extended Data Fig. 2 | Rate of methane production observed in long-term enrichment experiments. The rate at 25 weeks was derived from the value measured at 25 weeks minus the initial value (0) divided by the time elapsed in days. The rate at 60 weeks was derived from the difference between methane

measured at 60 weeks and methane measured at 25 weeks divided by the time elapsed since the 25-week measurement. Data are from single measurements owing to limited sample availability, except for the ASW blank ($n=3$), for which the mean is displayed. Source Data are available online.



Extended Data Fig. 3 | 16S rRNA iTAG composition and metabolic processes detected in metatranscriptomes. Top, taxonomic composition at family level or deeper (where possible) of 16S rRNA iTAG sequences for taxa present as more than 1% of the total abundance in at least one sample. Taxa present at an abundance of less than 1% in every sample were grouped into the 'Others' category. Bottom, an overview of categories of expressed genes in each sample

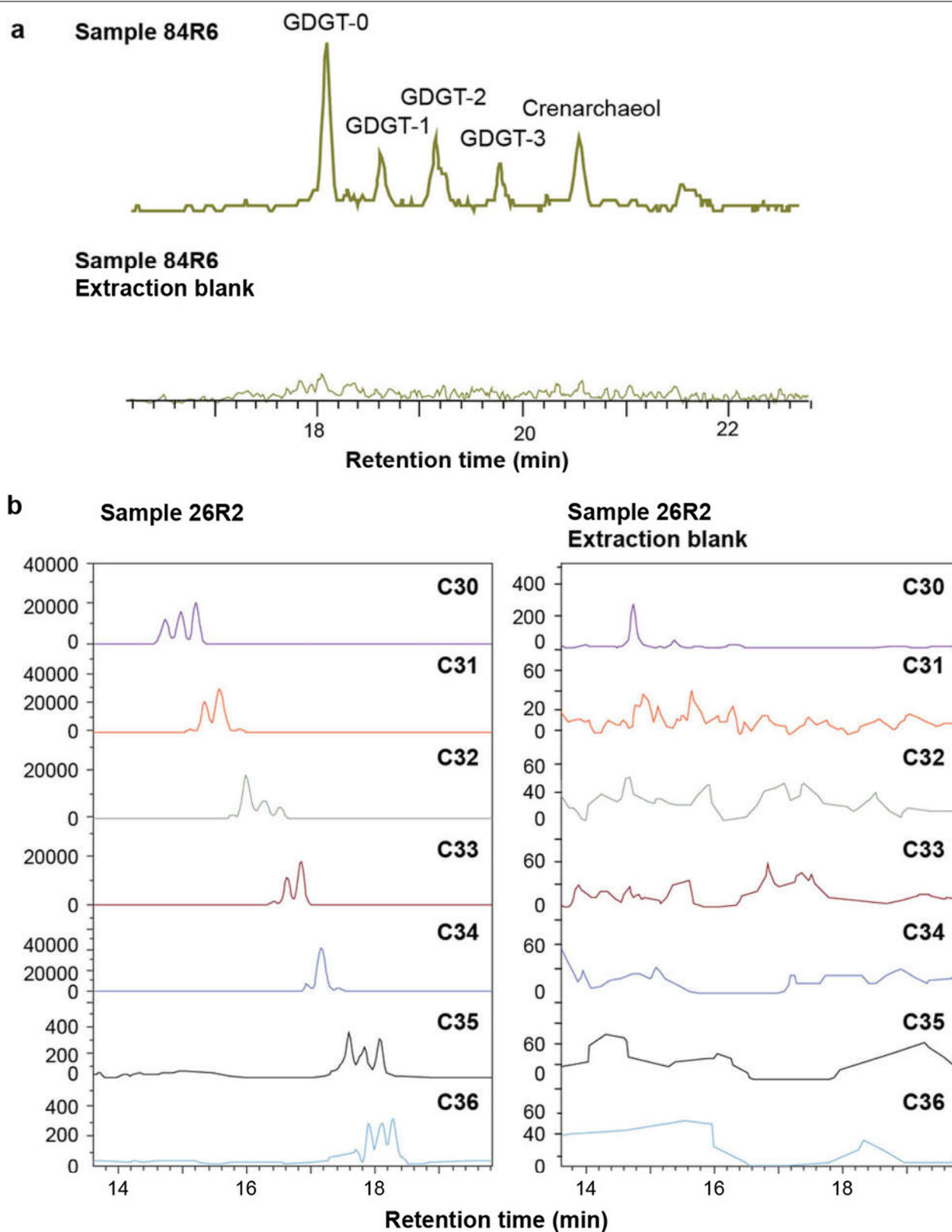
is given below the iTAG composition for each sample. The presence (black bars) or absence (white bars) indicates the detection or non-detection of genes associated with the processes and activities listed in that sample. A detailed discussion of metabolic pathways is provided in the Supplementary Information. Annotations at higher taxonomic resolution are available upon request from the corresponding author.



Extended Data Fig. 4 | See next page for caption.

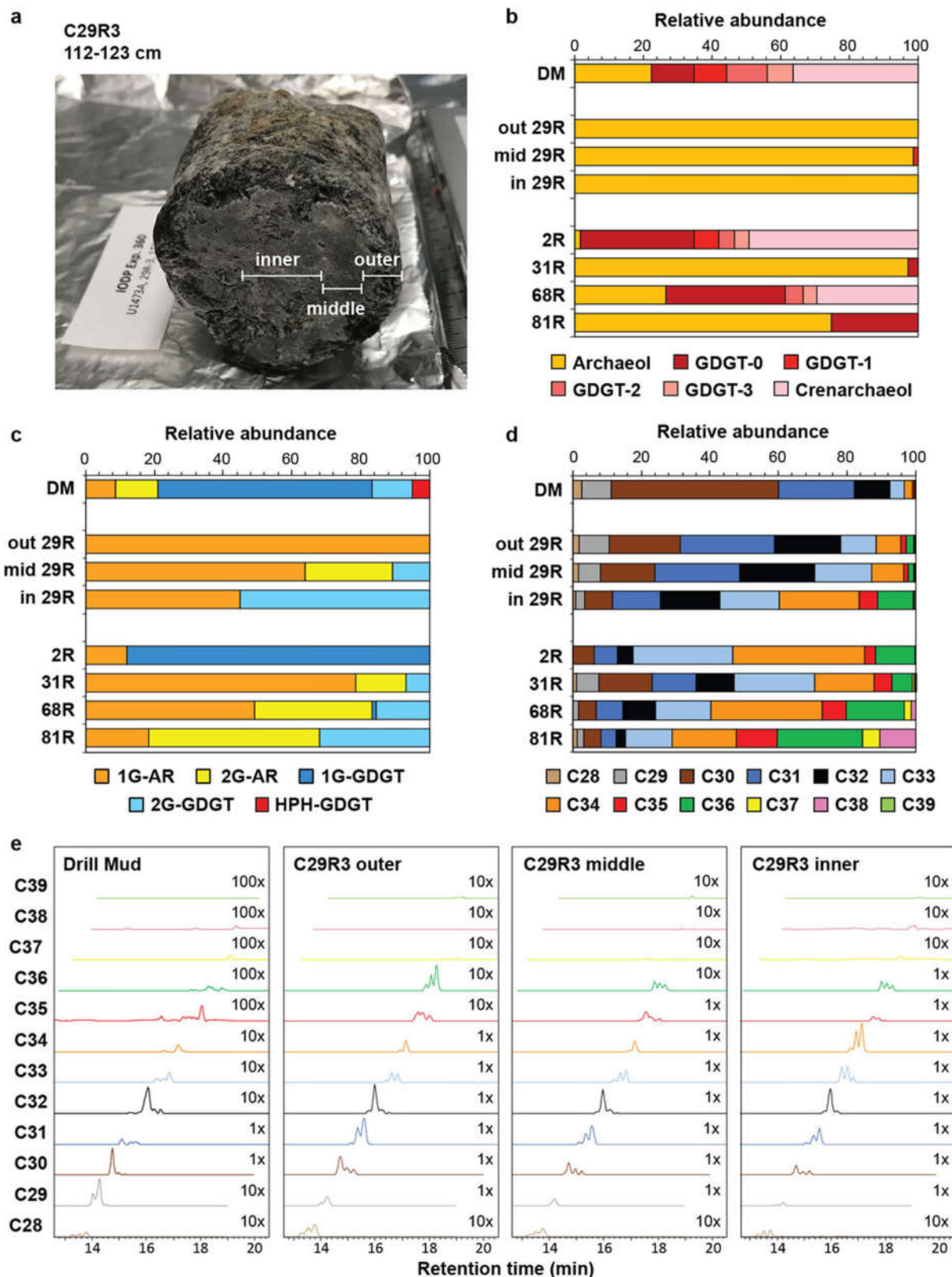
Extended Data Fig. 4 | Non-metric multidimensional scaling and clustering analyses of detected prokaryotic OTUs and transcripts. a. Non-metric multidimensional scaling analysis performed on the Jaccard distance matrix (Ward clustering using hclust; see Methods) of prokaryotic OTU presence or absence data for 11 biologically independent samples spanning the depth of hole U1473A showing only environmental vectors supported by $P < 0.05$. P values were generated as $P = N + 1/n + 1$ after a goodness-of-fit statistic, which is the squared correlation coefficient, and was calculated on 999 random permutations of the data using the vegan package of R. **b.** Clustering analysis of presence or absence prokaryotic OTU data for 11 samples spanning the depth of hole U1473A. Hierarchical clustering dendrogram based on a Jaccard

distance matrix of the presence or absence data for 99% OTUs from 11 IODP Expedition 360 samples. Jaccard similarity values of > 0.8 , calculated with the clusterboot function in R, suggest a stable cluster. Shading highlights several samples that share certain OTUs. From left to right (1–4): 1, samples share SAR11 clade II and SAR406; 2, samples share Nitrosopumilaceae; 3, samples share SAR11 clade II; and 4, samples both contain the lowest OTU counts. **c.** Clustering analysis of curated transcripts for 11 samples within the functional categories presented. Clustering analysis was based on $\log + 1$ -transformed FPKM values and the Ward method, and a distance matrix was constructed using the Manhattan method and the pvclust function in R.



Extended Data Fig. 5 | Lipid sample blanks. a, Representative UHPLC-APCI-MS chromatogram, showing the detection of core GDGTs in sample 84R6 and the sample extraction blank. **b**, Representative multiple-reaction-monitoring

HPLC-ESI-MS chromatograms for selected bacterial DEG lipids with summed chain lengths from C30 to C36 in sample 26R2 and the sample extraction blank. Lipid data are from single measurements owing to sample constraints.



Extended Data Fig. 6 | Drilling mud contamination control during lipid analysis. **a**, Depiction of outer, middle and inner sections subsampled from whole round core (WRC) sample 29R3 (259.03 mbsf) in order to test the influence of drilling mud (DM) on the lipid biomarker composition. **b–d**, Composition of archaeal core lipids (**b**), IPLs (**c**) and bacterial DEGs (**d**) in the drilling mud, subsampled WRC sample 29R3 and selected rock samples

analysed in this study. **e**, Extracted ion chromatograms of targeted DEG lipids from C28 to C39 summed carbon chain lengths in the drilling mud and subsampled WRC sample 29R3, showing the differences in relative abundance and isomer composition among these samples. Lipid data are from single measurements owing to sample constraints. 2R1, 10.7 mbsf; 31R1, 274.6 mbsf; 68R2, 619.6 mbsf; 81R2, 714.9 mbsf. Source Data are available online.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

For Raman analysis of thin section samples we used the LabSpec 6® Spectroscopy Suite Software (https://www.horiba.com/en_en/products/detail/action/show/Product/labspec-6-spectroscopy-suite-software-1843/) for data acquisition and background subtraction of acquired spectra. The processed spectra were then compared with the BIO-RAD KnowItAll® Raman Spectral Library databases for identification of organic substances and with the RRUFF (TM) Project Raman database for identification of minerals (Lafuente et al. (2015) (Lafuente B, Downs R T, Yang H, Stone N (2015) The power of databases: the RRUFF project. In: Highlights in Mineralogical Crystallography, T Armbruster and R M Danisi, eds. Berlin, Germany, W. De Gruyter, pp 1-30). For lipid analyses, Sciex Analyst 1.6.3 and Sciex MultiQuant 3.0.3 (AB Sciex LP, Concord, Canada) were used for triple quadrupole mass spectral data acquisition and data processing. Bruker Compass 1.9 and Bruker Data Analysis Version 4.4 (Bruker Daltonics, Bremen, Germany) were used for quadrupole time-of flight lipid data acquisition and processing.

Data analysis

The software R was used with the pvclust package for nMDS and clustering analysis of OTU data. Trimmomatic v. 0.32 was used to trim adapter sequences in transcript data, FastQC v. 0.11.7 and FASTX-toolkit v. 0.014 were used to quality check and trim paired reads in transcriptome libraries, Trinity v. 2.4.0 was used for assembly of transcriptomes, Bowtie v. 2.3.4.1 was used to align reads to our assembly and for de novo assembly of reads in our libraries, Bowtie aligner was used to align reads to assembled contigs, RSEM was used to estimate expression levels of reads in our libraries, TMM was used to perform cross-sample normalization and to generate a TMM-normalized expression matrix. The Trinotate suite (TransDecoder v. 3.0.1) was used to identify coding regions, BLASTx and BLASTp against UniProt, Swissprot (release 2018_02) and RefSeq nr were used to look for homologies between our reads and databases. HMMER v. 3.1b2 was used to identify conserved domains by searching against the Pfam v. 31.0 database. SignalP v. 4.1 and TMHMM 2.0c were used to predict signal peptides and transmembrane domains, RNAMMER v.1.2 was used to identify rRNA homologies. BLAST against the KEGG, COG, SEED and MetaCyc databases using MetaPathways v. 2.0 was used to interpret functions and pathways active in our transcriptome data. For our iTAG data Raw sequence reads were processed through Trim Galore [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/], FLASH (ccb.jhu.edu/software/FLASH/) and FASTX Toolkit [http://hannonlab.cshl.edu/fastx_toolkit/] for trimming and removal of low quality/short reads. Quality filtering included requiring a minimum average quality of 25 and rejection of paired reads less

than 250 nucleotides. Operational Taxonomic Unit (OTU) clusters were constructed at 99% similarity with the script pick_otus.py within the Quantitative Insights Into Microbial Ecology (QIIME) v.1.9.1 software and 'uclust'. The OTU data table was transformed to a presence/absence table and the Jaccard method was used to generate a distance matrix using the dist.binary() function in the R package ade4. A hierarchical clustering dendrogram was created using hclust() and the stability of the clusters was evaluated using the clusterboot() function in the fpc package in R with 500 iterations.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

iTAG data for the 11 samples as well as 13 different negative controls (drilling muds and fluids, seawater, kit controls) are deposited in GenBank under BioProject PRJNA497074. Results are presented at ~phylum level, however taxonomic assignments at finer resolution are available from the authors upon request. Raw reads for transcriptome data have also been deposited in GenBank SRA under BioProject PRJNA497074. Assemblies for curated portions of the data presented are available upon request to the authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples available for microbiological studies were selected by the Co-Chief Scientists of Expedition 360. They were limited to ~10-20cm long whole round rock sections from recovered cores. Replication was not possible because there was a single hole drilled for all studies conducted during Expedition 360, and only the 10-20 cm sections of 10m cores were allocated to microbiologists. Each of the individual 10-20cm core sections was processed to remove exterior rock and remaining material was subdivided for all analyses presented here.
Data exclusions	Outside of subtraction of molecular data matching any contamination controls (as noted in methods), no data were excluded. All data collected from analyzed samples is presented.
Replication	Cell counts were replicated by processing two replicate 1ml tubes of material from each core sample. Alkaline phosphatase activity measurements were conducted on triplicate technical replicates of material for those analyses and results were consistent. Methane production measurements are single measurements due to limited sample availability, except for the artificial seawater blank (n=3), for which the mean value is presented. All attempts at replication were successful. Sufficient material was not available for replicating other analyses. Lipid biomarker measurements were single measurements due to limited sample material. Raman spectral analysis presented in Figure 3a was made at two spots on this particular feature of interest, 10 independent times per spot, with similar results. In total, we performed >2000 analyses (with shorter integration times) on similar features in the same thin section sample with similar results.
Randomization	Eleven rock samples were selected from our collection for analyses presented here, purely on visual observations of the rock samples, prioritizing samples from the full depth of the hole drilled, and samples with evident veins and signs of alteration.
Blinding	Blinding was not applicable to this study of rock samples and associated control samples beyond the fact that all samples were number coded and treated identically throughout preparation and analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Recurrent interactions in local cortical circuits

<https://doi.org/10.1038/s41586-020-2062-x>

Received: 10 October 2016

Accepted: 14 January 2020

Published online: 4 March 2020

 Check for updates

Simon Peron^{1,4✉}, Ravi Pancholi², Bettina Voelcker², Jason D. Wittenbach¹,
H. Freyja Ólafsdóttir^{1,3,5}, Jeremy Freeman¹ & Karel Svoboda¹

Most cortical synapses are local and excitatory. Local recurrent circuits could implement amplification, allowing pattern completion and other computations^{1–4}. Cortical circuits contain subnetworks that consist of neurons with similar receptive fields and increased connectivity relative to the network average^{5,6}. Cortical neurons that encode different types of information are spatially intermingled and distributed over large brain volumes^{5–7}, and this complexity has hindered attempts to probe the function of these subnetworks by perturbing them individually⁸. Here we use computational modelling, optical recordings and manipulations to probe the function of recurrent coupling in layer 2/3 of the mouse vibrissal somatosensory cortex during active tactile discrimination. A neural circuit model of layer 2/3 revealed that recurrent excitation enhances sensory signals by amplification, but only for subnetworks with increased connectivity. Model networks with high amplification were sensitive to damage: loss of a few members of the subnetwork degraded stimulus encoding. We tested this prediction by mapping neuronal selectivity⁷ and photoablating^{9,10} neurons with specific selectivity. Ablation of a small proportion of layer 2/3 neurons (10–20, less than 5% of the total) representing touch markedly reduced responses in the spared touch representation, but not in other representations. Ablations most strongly affected neurons with stimulus responses that were similar to those of the ablated population, which is also consistent with network models. Recurrence among cortical neurons with similar selectivity therefore drives input-specific amplification during behaviour.

Two circuit motifs contribute to neural dynamics in cortical layer 2/3 (L2/3): recurrent excitation, which may cause amplification^{1–5,11,12}, and feedback inhibition, which may account for the sparse activity typical of L2/3^{12–14}. We explored the role of these motifs in an integrate-and-fire model of L2/3 of the mouse vibrissal somatosensory (‘barrel’) cortex constrained by measured physiological properties^{15,16} (Methods).

To model input-specific recurrent coupling, we restricted the sensory input to a subnetwork of the excitatory neurons (200 out of 1,700), corresponding to the number of L2/3 barrel cortex neurons that respond to active touch^{7,17} (Fig. 1a). We simulated touch-related input to L2/3 based on recordings of their L4 inputs¹⁸ (Methods). To measure how accurately neural activity reflects the sensory input, we computed an ‘encoding score’ (R_{stimulus}) by cross-correlating the spike rate of each neuron with the input (Methods).

We varied recurrence by changing the connection probability within the input-recipient subnetwork (subnetwork connectivity; synaptic conductance was scaled proportionately⁵) (Methods). For the connectivity of each subnetwork, we matched the distribution of the encoding score to that measured *in vivo*⁷ by adjusting the strength of the L4 input. Subnetworks with connectivity equal to and moderately increased

relative to the rest of the network (non-subnetwork connectivity, 0.2; subnetwork connectivity range, 0.2–0.4) produced responses that were consistent with those observed *in vivo*⁷ (Fig. 1b, Methods). The strength of sensory input required to match *in vivo* responses declined with increasing subnetwork connectivity. Amplification—defined as the ratio of network output to network input—therefore increased with subnetwork connectivity¹ (Fig. 1c). Additional increases in subnetwork connectivity (>0.4) produced all-or-none network responses, in which a transient input drove the network into a persistently active state¹⁹ (Extended Data Fig. 1).

Overall, subnetwork behaviour fell into three regimes, each of which produced a distinct response to removal (or ablation) of a small number of neurons. Subnetworks with low connectivity (0.2) amplified little, and were resistant to ablation (Fig. 1c–f). Encoding scores for spared neurons increased after simulated ablation of 10% of the subnetwork¹⁴, owing to reduced feedback inhibition^{13,20} (encoding score, from 0.237 ± 0.027 to 0.274 ± 0.032 (grand median \pm adjusted median absolute deviation (MAD)); $P < 0.001$, Wilcoxon signed-rank test, across $n = 30$ simulated networks with different randomized connectivity and initial conditions) (Fig. 1d–f, Methods).

¹Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. ²Center for Neural Science, New York University, New York, NY, USA. ³Department of Cell and Developmental Biology, University College London, London, UK. ⁴Present address: Center for Neural Science, New York University, New York, NY, USA. ⁵Present address: Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands. ✉e-mail: speron@nyu.edu

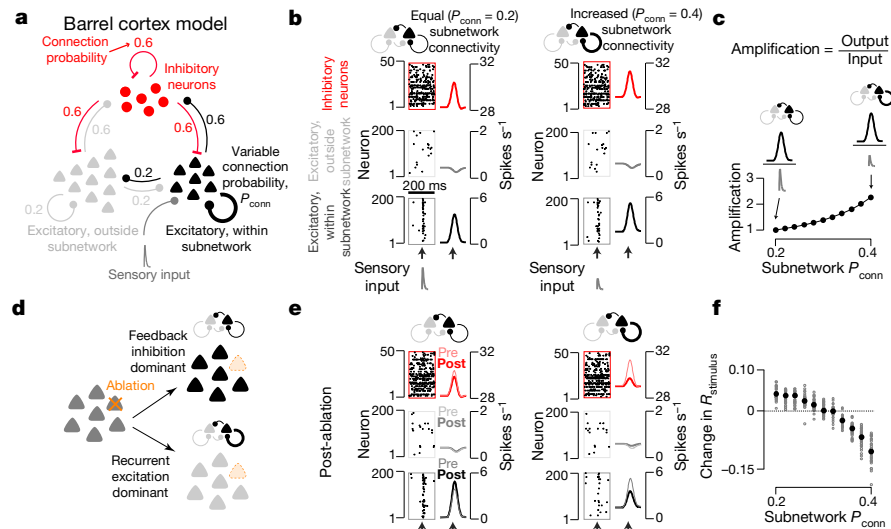


Fig. 1 | Ablation effect in simulated cortical L2/3 network depends on excitatory connectivity. **a**, The L2/3 network model comprises a subnetwork of 200 excitatory neurons (black) receiving sensory input (dark grey), a larger excitatory population (1,500 neurons; light grey) without sensory input, and 300 inhibitory neurons (red). All populations are interconnected (connection probabilities listed in figure) (Methods). The probability and strength of connections within the excitatory subnetwork was varied (P_{conn} , thick black loop) (Methods). **b**, Model network responses aligned to input (arrows). Left, subnetwork connectivity for excitatory subnetwork equal to overall connection probability. Right, increased connectivity. Raster plots show a subset of neurons from an example network. Peri-stimulus time histograms (PSTHs) show mean values across all neurons and networks ($n = 30$). Bottom,

excitatory neurons within subnetwork; middle, excitatory neurons outside subnetwork; top, inhibitory neurons. **c**, Amplification, defined as the ratio of network output to sensory input (Methods), as a function of subnetwork connectivity, normalized to $P_{\text{conn}} = 0.2$ (mean across 30 networks per P_{conn}). **d**, Predicted effects of ablation. In the equal-connectivity case (top), feedback inhibition dominates and responses among spared neurons increase. In the increased-connectivity case (bottom), recurrent excitation dominates and responses decline. **e**, As in **b**, but after ablation of the 20 neurons with the strongest encoding score. **f**, Effect of ablation on stimulus encoding, as a function of subnetwork connectivity. Grey points denote cross-neuron median for individual network. Black circles denote the grand median of 30 simulated networks.

Subnetworks with increased connectivity (for example, 0.4) exhibited stronger amplification¹⁴ (Fig. 1c). Ablations reduced encoding scores for spared neurons, from 0.243 ± 0.049 to 0.143 ± 0.028 (connectivity = 0.4; $P < 0.001$, $n = 30$ networks), because of reduced amplification (Fig. 1d–f). Networks in the all-or-none regime¹⁹ were robust to ablation, and maintained their all-or-none response (Extended Data Fig. 1). The response of subnetworks to ablations therefore distinguishes the three regimes.

We next performed a similar analysis for actual L2/3 networks during behaviour. We trained mice with a single spared whisker on an object localization task (Fig. 2a, b, Methods) and recorded neural activity with volumetric imaging⁷ in L2/3 of the barrel column corresponding to the spared whisker ($8,126 \pm 2,436$ (mean \pm s.d.) neurons per mouse, $n = 16$ mice) (Fig. 2c, Extended Data Table 1, Methods). Activity in a subset of neurons encoded whisker position (hereafter whisking neurons), whereas others responded to touch-induced changes in whisker curvature (hereafter touch neurons)⁷ (Fig. 2d, e). An encoding model generated a prediction of neural activity from vibrissal kinematics (Methods). Correlating the prediction with the actual neural activity yielded an encoding score, which was used to assign neurons to the touch and/or the whisking representations⁷ (Methods). Across 16 mice (Extended Data Table 1), 901 \pm 539 of the imaged neurons encoded touch (fraction: 0.108 ± 0.051) and 865 \pm 364 encoded whisking (fraction: 0.106 ± 0.028).

We probed the roles of recurrence by ablating members of the touch representation and examining the effect on spared neurons. Several excitatory neurons were ablated using multiphoton excitation^{9,10} (Extended Data Figs. 2, 3). Ablating a small proportion of strong touch cells (16.8 ± 12.8 neurons, 6% of touch neurons in the barrel column of the spared whisker, $n = 9$ mice; touch score, 70th \pm 35th percentile) (Fig. 2g) reduced responses to touch in the spared touch neurons (Fig. 2g, h). The touch-encoding score (R_{touch}) across touch neurons declined (from 0.123 ± 0.021 to 0.100 ± 0.037 (grand median \pm adjusted MAD); $n = 9$ mice, 8,392 neurons, $P = 0.004$, Wilcoxon signed-rank test

for mouse medians, paired by mouse) (Fig. 2j, Methods), as did the touch neuron count (from 932 ± 634 to 716 ± 469 (mean \pm s.d.), calculations exclude ablated neurons) (Methods). The whisking-encoding score (R_{whisking}) did not change (from 0.116 ± 0.013 to 0.115 ± 0.024 ; 6,975 neurons, $P = 0.820$; neuron count: from 775 ± 267 to 721 ± 241) (Fig. 2i, k).

In the model, more extensive ablations caused larger declines in encoding scores (Extended Data Fig. 4). In agreement with this prediction, the decline in touch representation increased as more of the touch representation was ablated (Pearson correlation of change in R_{touch} and net R_{touch} ablated, $R = -0.794$, $P < 0.001$ across all 24 ablations; $R = -0.779$, $P = 0.013$ for the 9 touch cell ablations) (Fig. 2l). Touch neurons proximal (15–35 μm) to the ablated cells experienced a larger decline in R_{touch} values than those distal (115–135 μm) from the ablated cells (Fig. 2m). The effects of ablating the touch neurons decayed over a distance ($\lambda = 87 \mu\text{m}$, exponential fit) (Fig. 2m) similar to the spatial scale of local recurrent connectivity in the rodent sensory cortex²¹. Whisking neurons exhibited no distance-dependent changes (Extended Data Fig. 5). The declining touch representation was not caused by changes in whisker movement or behaviour (Extended Data Fig. 6). This result is consistent with amplification of touch responses by recurrent excitation in L2/3.

By contrast, ablating a subset of strong whisking neurons (12.7 ± 5.7 neurons, approximately 4% of whisking neurons in the barrel column of the spared whisker, $n = 7$ mice; whisking score: 66th \pm 37th percentile) produced no effect on either the touch representation (R_{touch} from 0.112 ± 0.081 to 0.095 ± 0.069 , $n = 7$ mice, 5,866 neurons; $P = 0.109$, count from 838 ± 401 to 899 ± 548) or the whisking representation (R_{whisking} from 0.108 ± 0.076 to 0.107 ± 0.076 , $n = 6,161$; $P = 0.812$, count from 880 ± 387 to 974 ± 558) (Fig. 2j, Extended Data Fig. 7). Similarly, ablating silent neurons (event rate below 0.025 Hz; 16.3 ± 2.6 neurons, $n = 8$ mice) did not change the touch representation (R_{touch} from 0.115 ± 0.021 to 0.107 ± 0.023 , $n = 8$ mice, 7,110 neurons; $P = 0.383$, count from 889 ± 713 to 926 ± 681) (Fig. 2k, Extended Data Fig. 7, Methods) or the whisking representation (R_{whisking} from 0.115 ± 0.014 to 0.114 ± 0.015 , 7,684 neurons; $P = 0.844$, count from

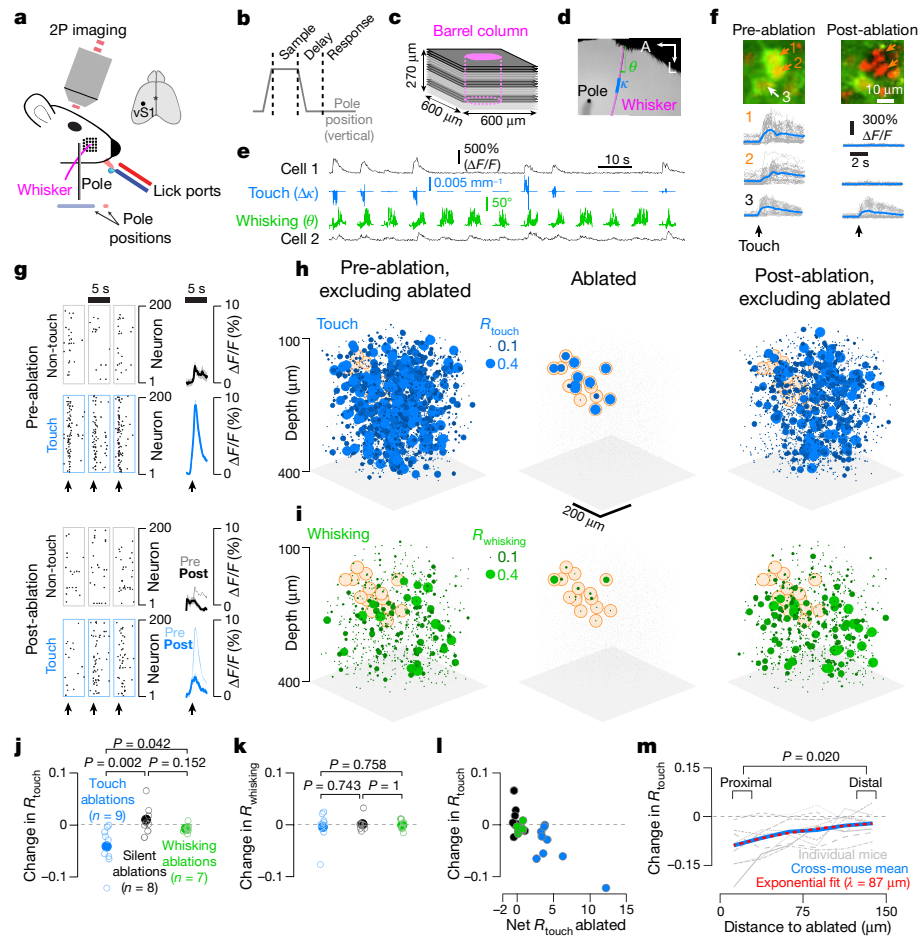


Fig. 2 | Touch networks of the barrel cortex are sensitive to ablation. **a**, Mice used one whisker to locate a pole and reported the perceived position by licking one of two lick ports (red or blue). **b**, Task epochs (Methods). **c**, Imaging planes were centred on the barrel column of the spared whisker (magenta). Three planes were imaged simultaneously (same shade of grey). **d**, Example video frame showing the whisker (magenta), pole (black), whisker curvature (κ , blue), and whisker angle (θ , green). **e**, Neuronal activity for a touch (cell 1) and whisking (cell 2) cell. Blue denotes change in whisker curvature ($\Delta\kappa$); green denotes whisker angle (θ). **f**, Ablations. Pyramidal neurons 1 and 2 (orange) were ablated (green, GCaMP6s; red, mCherry). Bottom, touch-aligned neuronal responses. **g–i**, Example experiment. **g**, Touch responses (dots, calcium events) (Methods) for a subset of neurons before (top) and after (bottom) ablation. Grey boxes denote 200 non-touch, non-whisking neurons; blue boxes denote 200 touch neurons. Right, touch-

961 ± 464 to 904 ± 384). Touch ablations produced significantly different changes in encoding scores among touch (but not whisking) representations than either whisking or silent ablations, whereas whisking ablations did not produce significantly different changes from silent ablations (Fig. 2j, k). Non-specific effects of ablation therefore do not contribute to changes in the touch representation after ablation.

Ablation of whisking neurons did not degrade the whisking representation. Our network model suggests that the slower kinetics of the whisking input do not account for this lack of effect, as networks having increased subnetwork connectivity (0.4) with slower input ($t_{\text{peak}} = 50$ ms, versus 10 ms for touch simulations) were sensitive to simulated ablation (Extended Data Fig. 8, Methods). Touch input is in-phase across the neural population^{17,18}, which engages recurrent excitation. By contrast, individual neurons encode whisking input with different phases²²; this asynchronous population response is expected to engage recurrent excitation less effectively. Therefore, the lack of sensitivity to ablation in the whisking population does not necessarily indicate an absence of recurrent coupling.

aligned mean $\Delta F/F$ values across these neurons for the strongest 5% of touches. Shading denotes s.e.m. **h**, Left, map for touch cells before ablation. Sphere size denotes R_{touch} values. Grey dots denote other neurons. Orange denotes the position of ablated neuron. Centre, R_{touch} values for ablated neurons. Right, R_{touch} values after ablation. **i**, As in **h**, but for whisking neurons. **j**, Median change in R_{touch} values across neurons in individual mice after ablation. Blue, touch ablations; black, silent neuron ablations; green, whisking ablations. P values determined by Wilcoxon rank-sum test. **k**, As in **j**, but for R_{whisking} values. **l**, Ablation effect on change in R_{touch} value as a function of R_{touch} values summed over the ablated neurons. Colour as in **j**. **m**, Distance-dependence (with respect to closest ablated neuron) of change in R_{touch} after ablation. Proximal change in R_{touch} : -0.132 ± 0.246 (grand median \pm adjusted MAD), distal: -0.056 ± 0.241 ; P value determined by Wilcoxon signed-rank test for proximal versus distal medians, paired by mouse, $n = 9$ mice.

In recurrent networks connected in a feature-specific manner⁵, the effects of ablation on spared neurons should increase with the similarity of their tuning to the ablated population^{23,24}. We tested this intuition in our model, defining the response similarity as the correlation of single neuron activity with the mean activity of the ablated neurons (Fig. 3a, Methods). In networks with increased subnetwork connectivity, neurons with high response similarity showed the largest decline in encoding score after ablation (Fig. 3b). In networks without increased subnetwork connectivity, the relationship disappeared (Fig. 3b).

We performed a similar analysis on our experimental data. Because not all neurons were recorded simultaneously (Fig. 2c, Methods), it was not possible to compute a mean across ablated neurons or correlations between that mean and the activity of individual neurons. We therefore devised a response similarity metric that did not require simultaneous recording. For each neuron, we averaged pre-ablation responses across both trial types (Fig. 3c). Concatenating these yielded the trial-averaged $\Delta F/F$ of the neuron (Fig. 3d). Computing the mean trial-averaged $\Delta F/F$ across all ablated neurons provided the ablated

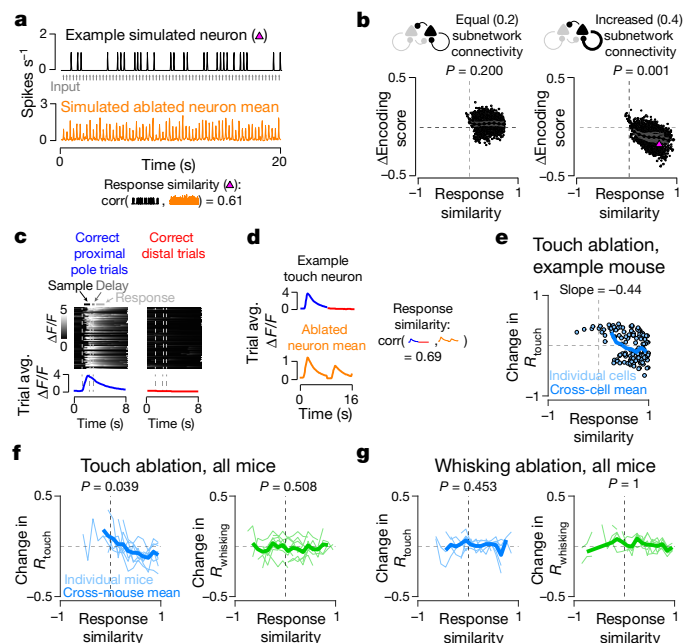


Fig. 3 | Effect of ablations depends on response similarity to ablated cells.

a, Response similarity in simulated networks. Top, example neuron spike rate (black). Grey arrows denote sensory input. Orange denotes mean spike rate across ablated neurons. Bottom, response similarity was computed by correlating neuronal spike rate with the mean ablated neuron spike rate. **b**, Dependence of the change in encoding score after ablation on response similarity in model networks. Grey dots denote individual neurons. Dark grey lines denote single network averages. P values for sign test that the slopes of linear fits across all networks ($n = 30$; $P_{\text{conn}} = 0.2$ case: 3,779 neurons; $P_{\text{conn}} = 0.4$ case: 3,466 neurons) are 0. Black line denotes grand mean across networks. Magenta triangle denotes neuron from **a**. **c**, Trial averaged response for example touch neuron. Heat maps show $\Delta F/F$ values for individual trials (proximal pole trials, left; distal, right). Bottom, trial-averaged $\Delta F/F$ values. **d**, Top, trial-averaged $\Delta F/F$ values for neuron in **c**. Bottom, the mean trial-averaged response across ablated neurons. Right, response similarity is the correlation of individual neuron trial-averaged response vectors with the mean across ablated neurons. **e**, Changes induced by ablation of touch neurons on the touch score as a function of response similarity in an example mouse. Blue circles denote individual neurons; blue line denotes binned mean (Methods). Slope for a linear fit of the points is given. **f**, Population data for all touch ablations for both touch (blue, left) and whisking (right, green) neurons. Thin coloured lines denote individual mouse mean values for a given response similarity bin; thick lines denote cross-mouse mean values. P values are from a sign test in which the slope of change in encoding score as a function of response similarity is 0 ($n = 9$ mice; 2,768 touch neurons; 1,085 whisking neurons). **g**, As in **f**, but for ablation of whisking neurons ($n = 7$ mice; 1,692 touch neurons; 935 whisking neurons).

neuron mean response. Response similarity was measured for each neuron as the correlation between its trial-averaged $\Delta F/F$ response and the mean response across all ablated neurons.

Changes in R_{touch} values after ablation depended on response similarity (Fig. 3e, f). R_{touch} values in neurons with high response similarity declined. Neurons with negative response similarity showed increased R_{touch} values, potentially owing to reduced feedback inhibition previously evoked by the ablated neurons²⁰ (Fig. 3e). Ablation of touch neurons had no effect on the whisking network (Fig. 3f). Ablation of whisker neurons had no effect on either representation (Fig. 3g).

Targeted photoablation allowed us to test the roles of recurrence in cortical circuits¹. The selective degradation of representations similar to the ablated neurons is consistent with amplification in recurrent networks^{1–4,11,12,23–26} (Fig. 1), but inconsistent with circuit models of sparse coding that are dominated by feedback inhibition^{12,14,20,27} or

models with all-or-none activity¹⁹. Our experiments reveal that cortical circuits can be surprisingly sensitive to damage targeting specific representations, despite remarkable robustness to other types of perturbation²⁸.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2062-x>.

- Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
- Chance, F. S., Nelson, S. B. & Abbott, L. F. Complex cells as cortically amplified simple cells. *Nat. Neurosci.* **2**, 277–282 (1999).
- Christie, I. K., Miller, P. & Van Hooser, S. D. Cortical amplification models of experience-dependent development of selective columns and response sparsification. *J. Neurophysiol.* **118**, 874–893 (2017).
- Miller, K. D. Canonical computations of cerebral cortex. *Curr. Opin. Neurobiol.* **37**, 75–84 (2016).
- Cossell, L. et al. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518**, 399–403 (2015).
- Lee, W. C. et al. Anatomy and function of an excitatory network in the visual cortex. *Nature* **532**, 370–374 (2016).
- Peron, S. P., Freeman, J., Iyer, V., Guo, C. & Svoboda, K. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron* **86**, 783–799 (2015).
- Peron, S. & Svoboda, K. From cudgel to scalpel: toward precise neural control with optogenetics. *Nat. Methods* **8**, 30–34 (2011).
- König, K., Becker, T. W., Fischer, P., Riemann, I. & Halbhuber, K. J. Pulse-length dependence of cellular response to intense near-infrared laser pulses in multiphoton microscopes. *Opt. Lett.* **24**, 113–115 (1999).
- Vladimirov, N. et al. Brain-wide circuit interrogation at the cellular level guided by online analysis of neuronal function. *Nat. Methods* **15**, 1117–1125 (2018).
- Rubin, D. B., Van Hooser, S. D. & Miller, K. D. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**, 402–417 (2015).
- Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51–58 (2013).
- Mateo, C. et al. In vivo optogenetic stimulation of neocortical excitatory neurons drives brain-state-dependent inhibition. *Curr. Biol.* **21**, 1593–1602 (2011).
- Barrett, D. G., Denève, S. & Machens, C. K. Optimal compensation for neuron loss. *eLife* **5**, e12454 (2016).
- Avermann, M., Tomm, C., Mateo, C., Gerstner, W. & Petersen, C. C. Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *J. Neurophysiol.* **107**, 3116–3134 (2012).
- Lefort, S., Tomm, C., Floyd Sarria, J. C. & Petersen, C. C. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* **61**, 301–316 (2009).
- Crochet, S., Poulet, J. F., Kremer, Y. & Petersen, C. C. Synaptic mechanisms underlying sparse coding of active touch. *Neuron* **69**, 1160–1175 (2011).
- Hires, S. A., Gutnisky, D. A., Yu, J., O'Connor, D. H. & Svoboda, K. Low-noise encoding of active touch by layer 4 in the somatosensory cortex. *eLife* **4**, (2015). <https://doi.org/10.7554/eLife.06619>
- Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**, 1498–1505 (2012).
- Chettih, S. N. & Harvey, C. D. Single-neuron perturbations reveal feature-specific competition in V1. *Nature* **567**, 334–340 (2019).
- Holmgren, C., Harkany, T., Svennenfors, B. & Zilberter, Y. Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol. (Lond.)* **551**, 139–153 (2003).
- Curtis, J. C. & Kleinfeld, D. Phase-to-rate transformations encode touch in cortical neurons of a scanning sensorimotor system. *Nat. Neurosci.* **12**, 492–501 (2009).
- Marshall, J. H. et al. Cortical layer-specific critical dynamics triggering perception. *Science* **365**, eaaw5202 (2019).
- Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A. & Yuste, R. Controlling visually guided behavior by holographic recalling of cortical ensembles. *Cell* **178**, 447–457 (2019).
- Lien, A. D. & Scanziani, M. Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.* **16**, 1315–1323 (2013).
- Daie, K., Svoboda, K. & Druckmann, S. Targeted photostimulation uncovers circuit motifs supporting short-term memory. Preprint at <https://www.biorxiv.org/content/10.1101/623785v1> (2019).
- Hansel, D. & van Vreeswijk, C. The mechanism of orientation selectivity in primary visual cortex without a functional map. *J. Neurosci.* **32**, 4049–4064 (2012).
- Li, N., Daie, K., Svoboda, K. & Druckmann, S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532**, 459–464 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Network model

We modelled the L2/3 network associated with a single barrel column—one of the most extensively studied cortical microcircuits^{29,30}—as a network of leaky integrate-and-fire neurons³¹. The dynamics of each neuron were governed by:

$$\tau \frac{dV_i}{dt} = V_i - V_r + R[I_i^{\text{exc}}(t) + I_i^{\text{inh}}(t) + I_i^{\text{ext}}(t)] \quad (1)$$

where V is the membrane potential, V_r is the rest/reset potential, τ is the membrane time constant, R is the input resistance, I_i^{exc} and I_i^{inh} are excitatory and inhibitory synaptic currents, I_i^{ext} is a current representing sensory stimulus drive (for example, from layer 4 inputs), and i indexes the neurons in the network. When the membrane potential reaches the spiking threshold, V_{th} , a spike is emitted, the membrane potential is reset to V_r , and the dynamics of the neuron are frozen for a short refractory period, t_{ref} . The synaptic currents follow kick-and-decay dynamics:

$$\tau_{\text{syn}} \frac{dI_i^{\text{syn}}}{dt} = -I_i^{\text{syn}} + \tau_{\text{syn}} \sum_{j,k} w_{ij} \delta(t - t_{jk}^{\text{syn}} - t_d) \quad (2)$$

where ‘syn’ denotes the type of synapse (either excitatory or inhibitor), τ_{syn} is the synaptic time constant, w_{ij} is a matrix of synaptic strengths from neuron j to neuron i , t_{jk} is the time of the k th spike of neuron j , and t_d is the spike transmission delay. The sum over j is over all neurons, while the sum over k is over all spikes from that neuron.

The network comprised 2,000 neurons, of which 1,700 (85%) were excitatory and 300 (15%) were inhibitory¹⁶. Excitatory neurons had $\tau = 30$ ms, whereas inhibitory neurons had faster dynamics with $\tau = 10$ ms^{15,16}. As the subthreshold dynamics for these neurons are linear, the behaviour of these neurons is invariant to changes of scale in V . The meaningful quantity is $\Delta V = V_{\text{th}} - V_r$, which we assume to be, on average, 35 mV for all neurons^{15,16}. Each neuron was assigned a value chosen uniformly within a range of $\pm 50\%$ of this mean value. Excitatory and inhibitory post-synaptic currents had time constants of $\tau_{\text{exc}} = 2$ ms and $\tau_{\text{inh}} = 3$ ms, respectively. The refractory period for spiking was $t_{\text{ref}} = 0.5$ ms. Synapses had a mean synaptic delay of $t_d = 0.6$ ms¹⁵; each individual synapse was assigned a value chosen within a range of $\pm 50\%$ of this mean value. Although all neural parameters could be modelled as random variables, we found that jittering ΔV for the neurons and t_d for the synapses provided sufficient heterogeneity to give a broad range of baseline spike rates across neurons and also prevented network synchronization. The input resistance, R , was factored into the synaptic weights and the magnitudes of the external currents, as described below.

The excitatory population was subdivided into two groups: a small input-recipient subnetwork of 200 neurons and the remainder of the excitatory population (1,500 neurons). Although all neurons in L2/3 probably receive touch-related input from L4³², only a small proportion are driven to spike after touch^{7,17}, justifying this model assumption. We neglected feedforward inhibition³³. This is justified because feedforward inhibition simply rescales input from L4², which in our model is adjusted to produce experimental observed population activity levels. Thus, each neuron belonged to one of three groups: the excitatory subnetwork (S), the remainder of the excitatory population (E), or the inhibitory population (I). Connections between neurons are determined by a block stochastic model. Given two neurons—the first from group A, the second from group B—there was a fixed probability of a connection from the first neuron to the second, denoted p_{AB} .

In our ‘equal subnetwork connectivity’ network, we used sparse connectivity between excitatory neurons, $p_{SS} = p_{SE} = p_{ES} = p_{EE} = 0.2$, and more dense connectivity both within the inhibitory population as well

as between the excitatory and inhibitory populations, $p_{II} = p_{IS} = p_{IE} = p_{SI} = p_{EI} = 0.6$ (refs. ^{15,16}). All connections had the same strength, which was chosen so that the resulting post-synaptic potential (PSP) is 1 mV (refs. ^{15,16}). Because it is the only connectivity parameter that was varied systematically, we denote p_{SS} as ‘ P_{conn} ’ in the rest of the text.

In the increased-subnetwork-connectivity version of the network, P_{conn} was increased to 0.4. To replicate the experimentally observed relationship between connectivity and synaptic strength⁵, the synaptic weight for these connections was increased to give PSPs of 1.6 mV. Networks with other levels of connectivity within the subnetwork were produced by linear interpolation or extrapolation of both P_{conn} and synaptic strength between the equal-subnetwork-connectivity and increased-subnetwork-connectivity cases.

The stimulus drive to the network was modelled as an external current targeting the excitatory subnetwork (group S; $I_i^{\text{ext}} = 0$ for all neurons in groups E and I). For each stimulus presentation, the waveform of the current was modelled with a beta distribution with shape parameters $\alpha = 3$ and $\beta = 5$. The beta distribution is defined on the interval $[0, 1]$, giving a distinct beginning and end to the stimulus. For the chosen shape parameters, the beta distribution has a value of 0 at its end points and a peak at $(\alpha - 1)/(\alpha + \beta - 2)$. To model the fast touch stimulus, the waveform was stretched in time so that the peak occurs 10 ms after the start (full-width half-maximum, 12.8 ms)¹⁸. The amplitude of the stimulus waveform was chosen so that the network response matched the experimental data. All neurons also receive tonic background input in the form of a Poisson spike train of excitatory spikes with a frequency of 5,000 Hz for excitatory neurons and 2,000 Hz for inhibitory neurons; these values were selected so that the tonic firing rate of these populations were approximately 0.5 Hz for excitatory neurons and 10 Hz for inhibitory neurons^{34–36}.

Simulations were performed in Python using the Brian2 simulation package³⁷ with a step-size of $dt = 0.1$ ms. For each randomly sampled network connectivity, the network was first simulated for 20 s of model time (corresponding to 66 stimulus presentations) and spike trains for all neurons were recorded. The activity of each neuron was then given an encoding score, which quantifies the signal-to-noise ratio of the representation of the stimulus: the spike train of the neuron was convolved with a Gaussian kernel (standard deviation, 20 ms) to produce a firing rate; the firing rate as well as the stimulus waveform was down-sampled by a factor of 5 (to a sample period of 0.5 ms) and the normalized cross-correlation between the signals was computed for leads/lags up to 10 ms; the peak value this cross-correlation is the encoding score.

We first ran exploratory simulations to constrain the strength of the input to the subnetwork to match physiological data⁷. We simulated both the equal-subnetwork-connectivity ($P_{\text{conn}} = 0.2$) and the increased-subnetwork-connectivity ($P_{\text{conn}} = 0.4$) networks across a range on sensory input strengths. In both cases, we examined the distributions of excitatory neuron encoding scores as a function of input strengths, selecting the input strength that produced a distribution most closely matching the experimental data (in terms of distribution shape and fraction of neurons encoding the stimulus). With input strengths defined for these two cases, we then used linear interpolation or extrapolation to select the input strength for other amounts of subnetwork connectivity. Because this procedure ensured a fixed network output following the simulated sensory stimulus, amplification (ratio of network output to sensory input) (Fig. 1c) was defined as the inverse of the sensory input strength². Amplification was normalized to the case where neurons within the input-recipient subnetwork had connectivity probabilities equal to the non-input recipient excitatory population (that is, $P_{\text{conn}} = 0.2$).

The final set of simulations explored the effects of targeted ablation across a range of subnetwork connectivity levels (Figs. 1, 3, Extended Data Fig. 1), number of ablations (Extended Data Fig. 4), and input kinetics (Extended Data Fig. 8). To simulate targeted ablation, we used

30 randomized network instances for each subnetwork connectivity level, number of ablations, and input rise time, and calculated the pre-ablation encoding score for each neuron. The 20 excitatory neurons with the highest encoding score were then removed (all outgoing synaptic strengths set to 0), the simulation was repeated, and the encoding scores after ablation were computed. Examination of the effects of the number of ablated neurons was done with the top 0, 10, 20, 50 and 100 neurons (Extended Data Fig. 4). This was repeated for 30 different realizations of the stochastic network connectivity. Neurons were considered to be part of the sensory input representation if they had an encoding score above 0.1; the effect of the ablation (Fig. 1e, f) was quantified as the change in encoding score across all neurons that met the 0.1 encoding score criteria either before and/or after ablations. The response similarity analysis (Fig. 3) used a more stringent encoding score cut off of 0.25; using a cut off of 0.1 did not alter the result, but reduced the magnitude of the observed effect. The ablated neurons were excluded for calculations of the distributions encoding scores before and after ablation, as well as in constructing PSTHs (Fig. 1b, e).

In modelling the effect of the number of neurons ablated on the change in encoding score (Extended Data Fig. 4), we restricted our modelling to the increased subnetwork connectivity case, $P_{\text{conn}} = 0.4$. For simulations of ‘whisking’ input (Extended Data Fig. 8), we shifted from a ‘touch’-like stimulus, peaking 10 ms after stimulus onset, to one peaking 50 ms after onset to mimic the response to whisking input observed in vivo^{17,18,22,35}. We examined both the equal ($P_{\text{conn}} = 0.2$) and increased ($P_{\text{conn}} = 0.4$) connectivity cases.

For large subnetwork connectivity ($P_{\text{conn}} > 0.4$), the network transitioned to all-or-none behaviour, where strong enough input can drive the subnetwork into a state of persistent increased firing¹⁹ (Extended Data Fig. 1). To reset the network after such a transition, we introduced a strong pulse of excitatory current to the inhibitory population 300 ms after the stimulus onset. In this regime, the extrapolation scheme for setting the stimulus strength did not apply, because spike rate was dictated by network properties and not the input strength. Instead, the input strength affected the reliability with which a stimulus would cause a transition to the state of increased firing. We found that choosing a stimulus strength that was twice that dictated by our extrapolation scheme produced a reasonable number of stimulus-encoding neurons. Thus, we used this criterion to set the input strength in this regime.

PSTHs were constructed by averaging the Gaussian-convolved (kernel standard deviation, 20 ms) responses of individual neurons, aligned to the sensory input.

Determining model synaptic weights

Model synapses were defined by kick-and-decay dynamics of the post-synaptic currents. We set the synaptic weights to produce a desired amplitude of the resulting unitary PSP. Here, we derive the relationship between the synaptic weight and the PSP size that allows us to do this.

Assume that a synaptic current starts at $I = 0$ when a single spike of weight w arrives at $t = 0$. In the absence of any other spikes, the subsequent time-course of the current is found by integrating equation (2):

$$I(t) = we^{-t/\tau_{\text{syn}}} \quad (3)$$

To determine the resulting behaviour of $V(t)$, make the assumption that the PSP evolves according to a difference of exponentials:

$$V(t) = V_0(e^{-t/\tau_1} - e^{-t/\tau_2}) \quad (4)$$

Differentiating equation (4) and using equation (3), one can show that this form does indeed solve equation (3) if we choose:

$$\tau_1 = \tau_m, \tau_2 = \tau_{\text{syn}}, \text{ and } V_0 = \frac{Rw}{\tau_m/\tau_s - 1} \quad (5)$$

The PSP size is the maximum value of $V(t)$, which we can compute as:

$$V_{\text{max}} = V_0 a^{(1-a^{-1})^{-1}} (a - 1) \quad (6)$$

in which $a = \tau_m/\tau_s$ is the ratio of the time constants. Comparing this to the expression for V_0 in equation (6) we find our desired relationship:

$$Rw = V_{\text{max}} a^{(1-a^{-1})^{-1}} \quad (7)$$

The linearity of the synaptic dynamics allows us to use equation (7) to avoid explicitly determining the value of R .

Mice

All procedures were performed in compliance with the Janelia Research Campus Institutional Animal Care and Use Committee and the New York University University Animal Welfare Committee. Two transgenic lines were used for these experiments, differentiated in Extended Data Table 1 by a ‘j’ or ‘n’ in mouse ID. Mice with IDs starting with letter ‘j’ consisted of mice expressing nuclear localized mCherry in a Cre-dependent manner (*R26-LSL-H2B-mCherry*⁷; JAX 023139) crossed with mice expressing Cre in cortical pyramidal neurons (*Emx1-IRES-cre*³⁸; JAX 005628). In cortical L2/3, these mice expressed nuclear mCherry only in nuclei of excitatory neurons. Several (9 or 12) injections (450 μm deep, 300 μm apart; beveled pipettes, World Precision Instruments; 20 nl each, at 10 nl min^{-1} with a custom-built microinjector) of AAV2/1-syn-GCaMP6s (UPenn AV-1-PV2824)³⁹ were made in barrel cortex (3.6 mm lateral, 1.5 mm posterior) of young adult mice⁷ (6–8 week). After viral injection, a titanium head bar was attached to the skull and the craniotomy was covered with a cranial window. Craniotomies were always over the left hemisphere. Mice with IDs starting with ‘n’ consisted of mice expressing GCaMP6s in a Cre-dependent manner (*Ai162*⁴⁰; JAX 031562), crossed with *Slc17a7-IRES2-cre*⁴⁰ (JAX 023527) to restrict expression to in pyramidal neurons. Surgeries for these mice did not include viral injections but were otherwise identical. With the exception of j258836, all mice were male.

Behaviour

Approximately one week after surgery, mice were trimmed to a single row of whiskers (typically the C row) and placed on water restriction⁴¹ (1 ml per day). Training commenced 5–7 days after restriction onset. Mice were trained on an object-localization task^{42,43} (Fig. 2a, b). If the pole appeared in a range of proximal positions, the mouse would be rewarded with a small water drop for licking the right lick port; pole presentation at the distal position would be rewarded upon licking the left lick port (Fig. 2a). Trials consisted of a 1–1.2-s sample epoch followed by a 0.5–1.2-s delay epoch after which a 50- or 100-ms 3.4 kHz auditory response cue signalled to the mouse to respond (Fig. 2b). To prevent premature licking, the lick port was brought into tongue range by a motor (Zaber) only during the response epoch (Fig. 2b). Mice exhibiting excessive premature licking (licks before reward cue on >20% of trials; licks monitored with a laser beam; Thorlabs) were not used.

Mice were trimmed to a single whisker after reaching criterion performance ($d' > 1.5$ for two consecutive days). The spared whisker barrel column was identified using the neuropil signal, as described previously⁷. Whisker videography was performed at 400–500 Hz. Whiskers were tracked using an automated software pipeline⁴⁴ and then curated using custom browser software⁷.

Mice were assigned randomly to experimental groups (ablation type). Experimenters were not blinded to the group.

Imaging

Calcium imaging was performed using a custom two-photon microscope (<http://openwiki.janelia.org/wiki/display/shareddesigns/MIMMS>) with a 16 \times , 0.8 NA objective (Nikon)⁷. GCaMP (BG22; Chroma) and mCherry (675/70 filter; Chroma) fluorescence was imaged using

GaAsP PMTs (Hamamatsu). The 940 or 1,000 nm (Coherent) imaging beam was steered with a 16 kHz line rate resonant galvanometer (Thorlabs); a piezo collar (Physik Instrumente) moved the focus axially. The 512×512 pixel images were collected at 7 Hz, with three 600×600 - μm images per piezo cycle. Planes were spaced 15 μm apart. Scanimage⁴⁵ (Vidrio Technologies, <http://www.vidriotech.com>) controlled the microscope. Three planes, constituting a ‘subvolume’, were imaged simultaneously (4–6 subvolumes, spanning 45 μm each, 12–18 total planes, 180–270 μm in depth total), and power was modulated with depth using a length constant of 250 μm . Deeper subvolumes were typically imaged with higher power, using a similar length constant. Individual subvolumes were typically imaged for 50 trials per day, with all subvolumes usually visited on any given day. Alignment across days was performed as previously described^{7,46}.

Imaging data were processed using a semi-automated software pipeline that included image registration, segmentation, neuropil subtraction, $\Delta F/F$ computation, and event detection⁷. The de-noised $\Delta F/F$ trace, which consisted of the event amplitude trace convolved with event-specific exponential rise and decay, was used for analysis.

Neuronal classification

Neurons were classified using a linear-nonlinear encoding model^{7,47}. The model consisted of a cascaded generalized linear model that used a temporal and stimulus domain kernel to predict the activity of individual neurons given whisker angle, θ , or whisker curvature, κ , assuming Gaussian noise with input nonlinearities.

The model predicted the neuronal response, r (that is, $\Delta F/F$), as

$$r \sim \text{Norm}(z, \sigma^2)$$

$$z = f_1(s_1) \times k_1 + f_2(s_2) \times k_2$$

Here, s_1 and s_2 are the whisker angle, θ , and whisker curvature, κ , respectively. The terms f_1 and f_2 are static, point-wise nonlinearities comprising a weighted sum of 16 triangular basis functions

$$f = \sum_{i=1}^{16} w_i b_i(x)$$

in which x is the input (s_1 or s_2), with each b_i given by

$$b_i = \begin{cases} (x - x_{i-1}) / (x_i - x_{i-1}), & i > 1, x_{i-1} < x < x_i \\ (x_{i+1} - x) / (x_{i+1} - x_i), & i < N, x_i \leq x < x_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

k_1 and k_2 are temporal kernels consisting of 14 time points (2 s).

Thus, the model gives a z-scored prediction of neural activity, z , by fitting parameters k_1 , k_2 , f_1 and f_2 for given whisker kinematic parameters, s_1 and s_2 .

The model parameters k_1 , k_2 , f_1 and f_2 were fit using maximum likelihood with block coordinate descent. This procedure reliably estimated model parameters within three to five iterations.

To avoid degeneracy associated with arbitrary scaling factors on either the kernels or the nonlinearities, the nonlinearities were forced to have minimum of 0 and maximum of 1. Temporal kernels were unconstrained.

A prior was used to ensure smoothness of both the temporal kernels and the nonlinearities and prevent over-fitting. Specifically, the prior added a factor to the objective function penalizing excessive second derivatives of the temporal kernels and nonlinearities. Employing such a prior corresponds to maximizing the log-posterior, with the prior adding a small penalty to the objective function. To fit several thousand cells efficiently, the scale factor associated with this penalty was determined from a cross-validated inspection of a random subset of neurons.

The model was fit using fivefold cross-validation across trials. That is, a randomly selected 80% of trials were used for fitting, and 20% were used for evaluation, with 5 distinct evaluation groups per fit ensuring all data was used for fitting in exactly one case. The Pearson correlation between the predicted and actual $\Delta F/F$ yielded a measure of the quality of the model fit. This correlation was used as the encoding score for barrel cortex data: R_{touch} , based on $\Delta\kappa$ for touch, and R_{whisking} , based on whisker θ for whisking (Fig. 2d, e). A neuron was considered part of a representation if R_{touch} or R_{whisking} exceeded 0.1 and if the neuron score was above the 99th percentile of scores measured from matched shuffled neural activity. These criteria are more stringent than those used previously⁷, because ablation predominantly impacted neurons with high encoding scores. Using a less stringent criterion did not change the underlying conclusions, but did dilute the magnitude of the effect. For the response similarity analysis (Fig. 3), a threshold of 0.25 was used for the encoding score.

Shuffled activity was generated by randomizing the timing of the calcium events to construct a novel de-noised $\Delta F/F$ trace. Matched shuffled activity was selected by using neurons from the same imaging subvolume (that is, concurrently imaged to ensure identical mouse behaviour) sharing a similar event rate. Event rates were matched by partitioning the neurons from a subvolume into 10 equally sized bins (in terms of neuron count). Thus, in addition to the aforementioned encoding score threshold, a given neuron R value had to exceed the 99th percentile of R values obtained across neurons in the same subvolume and event rate bin whose event times were shuffled.

We examined robustness by partitioning individual days into two interdigitated pseudo-sessions, and measuring the correlation between encoding scores for the two pseudo-sessions⁷. The resulting correlations ranged between approximately 0.5 and 0.75. Neural classification (whisking and touch) was stable across days for trained mice. Furthermore, the touch neuron curvature kernels assumed ‘V’- or ‘L’-like shapes⁷, meaning that high magnitude curvature changes drove the largest responses, a result consistent with the known responses of these neurons. Finally, the kernels were stable across days⁷. Thus, changes in barrel cortex encoding after ablation are not a reflection of the variability inherent to the encoding model, but rather reflect genuine changes in the underlying representations.

Ablated neurons were always excluded from analysis of experimental data, including calculations of pre-ablation population encoding scores. Analyses involving data before and after ablation were pooled across several behavioural sessions. Pre- and post-ablation data each consisted of at least two (but typically three) behavioural sessions, with each subvolume sampled on most sessions. Given that a single subvolume was imaged for around 50 trials in a session, classification typically used around 150 behavioural trials per neuron (minimum for encoding model: 100 trials). Neurons participating in both representations in a given area were excluded from analysis. In all cases in which comparisons of pre- and post-ablation distributions were made, neurons were included for analysis if they met the criteria for representation membership (described above) in the pre- or post-ablation period, or during both periods.

Multiphoton ablation

Ablations^{9,48,49} were performed with 880 nm (Chameleon Ultra 2; Coherent) or 1,040 nm (Fidelity HP; Coherent) femtosecond laser pulses delivered through a 0.8 NA, 16 \times objective (Nikon) in mice that were awake but not performing the task or lightly anesthetized using isoflurane (1–2%). On the day of ablation, a new image was acquired and a warp field transform was used to find the target neurons⁴⁶. Imaging was not performed on the day of ablation; post-ablation analyses used the imaging data from the 2–4 behavioural sessions on the days after ablation. Typically, experiments consisted of three days of pre-ablation data collection, the ablation day, and three days of post-ablation data collection.

In virally transfected mice, the mCherry signal was used to restrict ablations to pyramidal neurons⁷; in mice endogenously expressing GCaMP, presence of fluorophore was used for this purpose. The neurons with strongest touch or whisking encoding scores within the spared barrel column were targeted for ablation. The percentage of touch and whisking neurons ablated was calculated in relation to the estimated 1,691 pyramidal neurons in L2/3¹⁶, of which 17% (287) belonged to each representation⁷. Neurons participating in both representations (touch and whisking) were avoided. For silent cell ablation, neurons with a calcium event rate below 0.025 Hz were targeted for ablation.

In all cases, approximately 2/3 of ablations were successful (Extended Data Fig. 2c). Proximity to vasculature, low baseline fluorescence, and excessive depth accounted for most failures. Thus, despite targeting the strongest neurons in a given representation, the actual representation strength of ablated neurons varied. The encoding score percentile of the ablated neurons is therefore reported in the text.

After ablation we consistently observed an increase in GCaMP6 fluorescence in the targeted neuron (Extended Data Fig. 2a, b). Taking advantage of this signature, the ablation protocol consisted of interleaved 'ablation' and 'evaluation' epochs (ablation duration: 50–200 ms; evaluation: 0.1–2 s, with longer evaluation times proving more reliable) (Extended Data Fig. 2b). Ablation epoch power started at the evaluation epoch power (25–100 mW, measured at the specimen) and rose linearly as necessary (up to 1 W) over the course of several seconds. The beam was focused on the brightest part of the targeted neuron using a pair of galvanometers (Cambridge Technology) and oscillated over a path spanning 1–2 μm . Evaluation epoch fluorescence data were collected using standard resonant galvanometer imaging⁷, although restricted to a single plane (approximately 30 Hz). Ablations were terminated after observation of the fluorescence rise in the target neuron.

Consistent with similar protocols^{10,50}, ablation did not produce off-target damage: calcium event rates for neurons adjacent (10–25 μm) to ablated cells did not change after the ablation of silent neurons (Extended Data Fig. 3a), and glial immunoreactivity was confined to the site of the lesioned neuron (Extended Data Fig. 3b–d).

Typically, 10–50 ablations were performed over the course of one hour. For histological analysis (Extended Data Fig. 3b–d), perfusion was performed 24 h after ablation in two *Emx1-IRES-cre* \times *LSL-H2B-mCherry* mice (9 and 28 ablations were successful in these mice, approximating typical experimental conditions). Alternating cryomicrotome (Leica) sections were exposed to antibodies for either the microglial marker IBA1 (Abcam, ab5076) or the astrocytic marker GFAP (Abcam, ab7260). Glial reactions were measured by first locating the centre of the ablated neuron in the glial immunoreaction image. An edge detection algorithm³⁹ that operated on an intensity image in angle-distance space from the neuron centre detected the extent of the reaction (Extended Data Fig. 3d). Specifically, intensity profiles were measured across a range of angles emanating from the point within the ablation. To delimit the glial reaction, large drops in intensity were detected. Glia was considered reactive if the intensity inside the detected reaction area was two standard deviations above background image intensity.

Following ablation, nearby neurons retained sensory responses (Fig. 2f), event rate (Extended Data Fig. 3a), and maintained structural integrity (Extended Data Fig. 3b, c, e). In a few instances ($n = 5$) the ablation termination protocol failed, resulting in more extensive lesions (Extended Data Fig. 3f). These experiments were excluded from the study.

We excluded ablated neurons as well as neurons within a cylinder centred on ablated neurons having a radius of 10 μm and a height of 60 μm from all analyses. This excluded neurons abutting ablated cells, and ensured no neurons within a typical glial reaction radius would be included.

Response similarity

For simulated data, response similarity was measured by taking the Pearson correlation of the Gaussian-convolved (kernel standard deviation, 20 ms) activity of an individual neuron with the mean Gaussian-convolved activity of the ablated neurons (Fig. 3a). For experimental data, response similarity was measured by correlating the individual neuronal trial averaged $\Delta F/F$ to the ablated neuron trial average mean. For each neuron, trial averaged $\Delta F/F$ was calculated by taking the mean $\Delta F/F$ across all correct proximal and distal pole trials (Fig. 3c), then concatenating these two vectors (Fig. 3d). The mean of these vectors across the ablated neurons constituted the ablated neuron mean (Fig. 3d). Response similarity is simply the Pearson correlation of the individual neuron trial averaged $\Delta F/F$ with the mean trial averaged $\Delta F/F$ across all ablated neurons. Single-network (Fig. 3b) or mouse (Fig. 3e–g) averages were computed with response similarity bins having a width of 0.1. The grand mean of these is shown as a dark line on these plots. Only strongly responding neurons (encoding score > 0.25) were considered in this analysis (other analyses use a cut-off value of 0.1).

Trial-averaged $\Delta F/F$ correlations were used instead of raw activity correlations because neurons were not all imaged simultaneously; only neurons in a given subvolume were imaged simultaneously. Because ablated neurons came from multiple subvolumes, response similarity had to use trial-averaged responses to allow for comparison across disjointly recorded populations.

Statistical analyses

Most statistical comparisons were performed using the Wilcoxon signed-rank test comparing paired medians within individual mice for two conditions (for example, before and after ablation, or proximal and distal encoding score change). For cases in which values had no natural pairing (comparison of different ablation types), the Wilcoxon rank-sum test comparing medians was used to compare distributions. To test whether encoding score change depended on response similarity (Fig. 3), we first fit a line to individual networks or mice (for example, Fig. 3e; the linear fit is distinct from the cross-cell mean that is shown). Next, we tested whether the slopes thus obtained were distinct from 0 across all networks or mice using the non-parametric sign test.

In all cases, we used the median of single-neuron values within a mouse. That is, we treated mice, and never neurons, as independent observations. Where relevant, the total number of neurons included across all mice was given. Where given, adjusted MAD was calculated by multiplying the median absolute deviation by 1.4826 so as to approximate the standard deviation under conditions of normality. Sample sizes were similar to those used by others in the field. No statistical tests were used to determine sample sizes.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data can be found at CRCNS (<http://crcns.org/>) at <https://doi.org/10.6080/K0Z31WWG>.

Code availability

Code for the simulations can be found at <https://github.com/jwittenbach/ablation-sim>. Code used for data analysis can be found at <https://github.com/peronlab/ablation>.

29. Petersen, C. C. & Crochet, S. Synaptic computation and sensory processing in neocortical layer 2/3. *Neuron* **78**, 28–48 (2013).

30. Petersen, C. C. H. Sensorimotor processing in the rodent barrel cortex. *Nat. Rev. Neurosci.* **20**, 533–546 (2019).

31. Brette, R. et al. Simulation of networks of spiking neurons: a review of tools and strategies. *J. Comput. Neurosci.* **23**, 349–398 (2007).
32. Bureau, I., von Saint Paul, F. & Svoboda, K. Interdigitated paralemniscal and lemniscal pathways in the mouse barrel cortex. *PLoS Biol.* **4**, e382 (2006).
33. Xue, M., Atallah, B. V. & Scanziani, M. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* **511**, 596–600 (2014).
34. Gentet, L. J. et al. Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nat. Neurosci.* **15**, 607–612 (2012).
35. Yu, J., Gutnisky, D. A., Hires, S. A. & Svoboda, K. Layer 4 fast-spiking interneurons filter thalamocortical signals during active somatosensation. *Nat. Neurosci.* **19**, 1647–1657 (2016).
36. Yu, J., Hu, H., Agmon, A. & Svoboda, K. Recruitment of GABAergic interneurons in the barrel cortex during active tactile behavior. *Neuron* **104**, 412–427 (2019).
37. Stimberg, M., Brette, R. & Goodman, D. F. Brian 2, an intuitive and efficient neural simulator. *eLife* **8**, e47314 (2019).
38. Gorski, J. A. et al. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
39. Chen, T. W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
40. Daigle, T. L. et al. A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480 (2018).
41. Guo, Z. V. et al. Procedures for behavioral experiments in head-fixed mice. *PLoS ONE* **9**, e88678 (2014).
42. O'Connor, D. H. et al. Vibrissa-based object localization in head-fixed mice. *J. Neurosci.* **30**, 1947–1967 (2010).
43. Guo, Z. V. et al. Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81**, 179–194 (2014).
44. Clack, N. G. et al. Automated tracking of whiskers in videos of head fixed rodents. *PLoS Comput. Biol.* **8**, e1002591 (2012).
45. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
46. Huber, D. et al. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**, 473–478 (2012).
47. Ahrens, M. B., Paninski, L. & Sahani, M. Inferring input nonlinearities in neural encoding models. *Network* **19**, 35–67 (2008).
48. Orger, M. B., Kampff, A. R., Severi, K. E., Bollmann, J. H. & Engert, F. Control of visually guided behavior by distinct populations of spinal projection neurons. *Nat. Neurosci.* **11**, 327–333 (2008).
49. Allegra Mascaró, A. L., Sacconi, L. & Pavone, F. S. Multi-photon nanosurgery in live brain. *Front. Neuroenergetics* **2**, 21 (2010).
50. Canty, A. J. et al. In-vivo single neuron axotomy triggers axon regeneration to restore synaptic density in specific cortical circuits. *Nat. Commun.* **4**, 2038 (2013).

Acknowledgements We thank S. Druckmann, S. Romani, D. Gutnisky, N. Li, J. Yu, H. Inagaki, N. Sofroniew and M. Economo for comments on the manuscript, A. Hu for histology, and H. Zeng for the *Ai162* transgenic mice. Funding was provided by the Howard Hughes Medical Institute. R.P. was supported by the National Institutes of Health (NIH) T32GM007308.

Author contributions S.P. and K.S. conceived the project. S.P., R.P. and B.V. performed the experiments, with assistance from H.F.O. J.D.W. performed the modelling, with input from J.F., S.P. and K.S. S.P., R.P., B.V., J.D.W., J.F. and K.S. analysed data and wrote the paper.

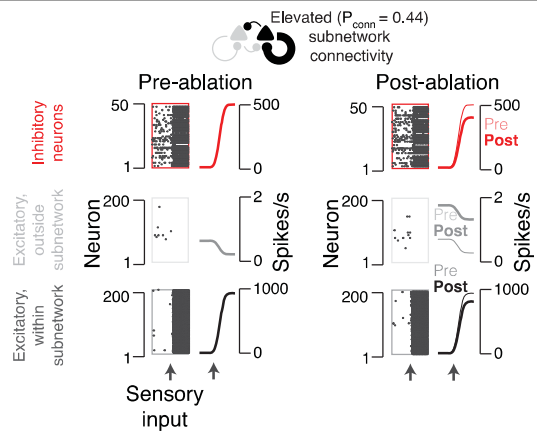
Competing interests The authors declare no competing interests.

Additional information

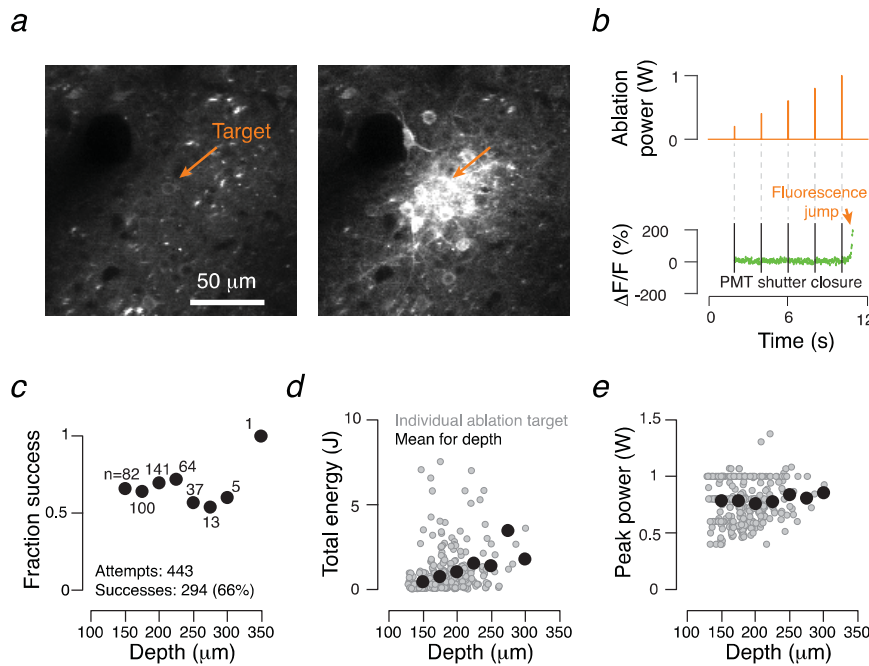
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2062-x>.

Correspondence and requests for materials should be addressed to S.P.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

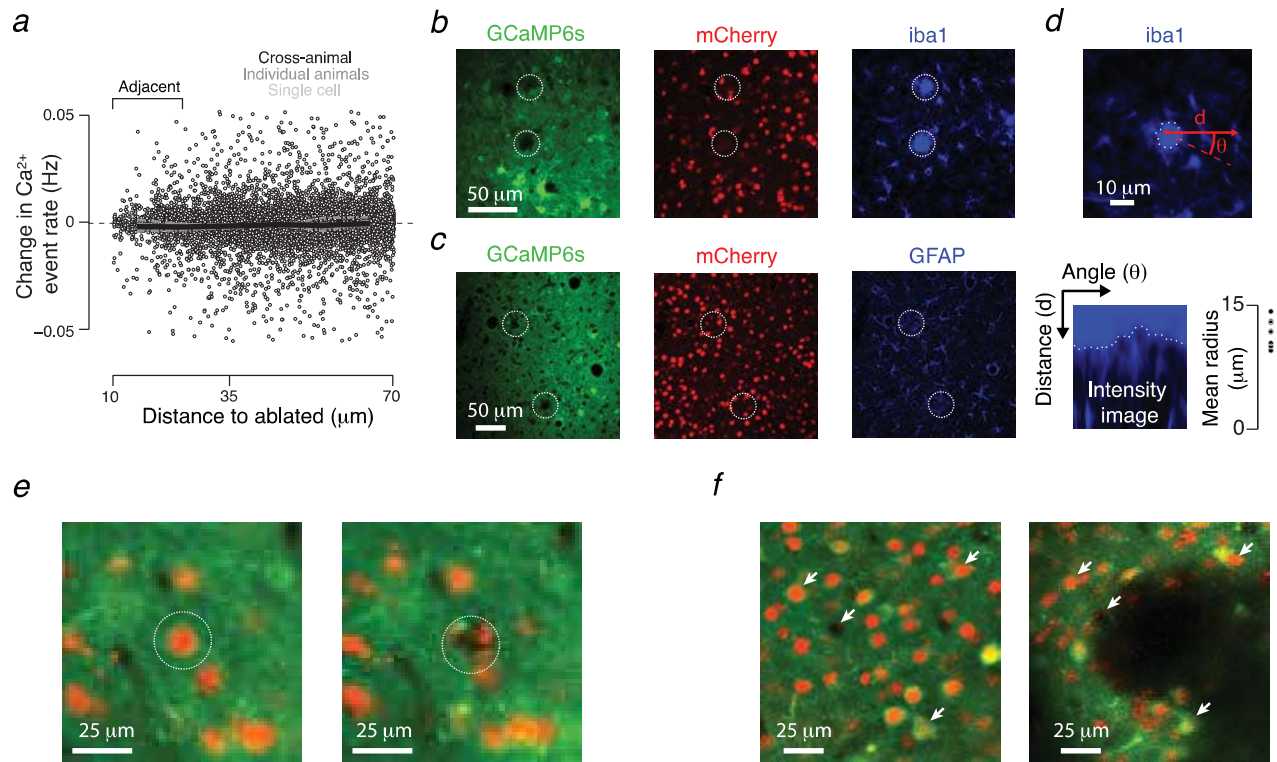


Extended Data Fig. 1 | Effect of ablation in a simulated network with hyper-connectivity (that is, $P_{\text{conn}} > 0.4$; here, $P_{\text{conn}} = 0.44$). Model network responses aligned to input (arrow, bottom). Raster plots show a subset of neurons from an example network. PSTHs show averages across all neurons and networks. Bottom, excitatory neurons within subnetwork. Middle, excitatory neurons outside subnetwork. Top, inhibitory neurons. Left, network response before ablation of 10% of the subnetwork neurons; right, response after ablation.



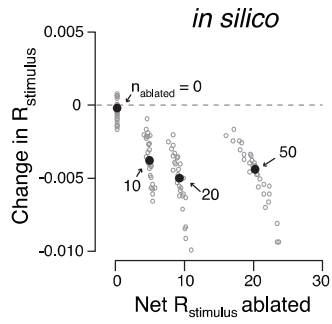
Extended Data Fig. 2 | Multi-photon ablation protocol. **a**, Example field of view immediately before (left) and after (right) ablation. Orange arrow denotes target neuron. **b**, Ablation protocol. Power (orange, top trace) during ablation epochs (100 ms; increased power, orange, top) gradually increased, and the PMT shutter was closed (black bars). During the intervening evaluation epochs, power was lower and constant (orange, top), and the PMT shutter was open. Ablation was terminated when GCaMP fluorescence at the target neuron (green) jumped (orange arrow). **c**, Success of ablation as a function of neuron

depth for all experiments included in this dataset. Individual points give mean success rate for given depth bin; bin size, 25 μm . **d**, Depth dependence of total energy deposition for successful ablations (successfully ablated cells only: $n = 293$ cells across 22 sessions, 14 mice; ablations from j250220 and j257218, along with 7 additional ablations from other mice were excluded owing to incomplete logging). Grey dots denote individual ablations. Black dots, means across 25- μm bins. **e**, As in **d**, but for peak power needed for ablation ($n = 293$ cells across 22 sessions, 14 mice).

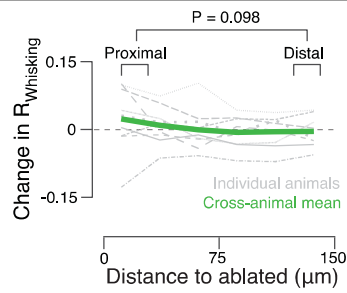


Extended Data Fig. 3 | Multi-photon ablation produces spatially localized effects. **a**, Change in calcium event rate (Methods) as a function of minimal distance to an ablated neuron after silent cell ablation. Individual neurons appear as grey points, with dark grey dashed lines showing single mouse averages and the dark black line showing the cross-mouse ($n = 8$ silent ablation mice) average. Event rates among neurons adjacent to ablated silent cells did not change (event rate before ablation: 0.014 ± 0.008 Hz, grand median \pm adjusted MAD; after ablation: 0.014 ± 0.007 Hz; $P = 0.055$ before versus after ablation, Wilcoxon signed-rank test, paired medians across $n = 8$ mice; 1,028 neurons across all mice) (Methods). **b**, Confocal ex vivo image from a mouse perfused 24 h after ablation. Ablation sites are indicated with dashed white circles. Green, GCaMP6s fluorescence; red, mCherry fluorescence; blue, microglial antibody IBA1 fluorescence. **c**, As in **b**, but blue shows immunoreactivity for the astrocytic marker GFAP. **d**, The spatial extent of glial reaction was measured by detecting the fastest intensity decline ridge

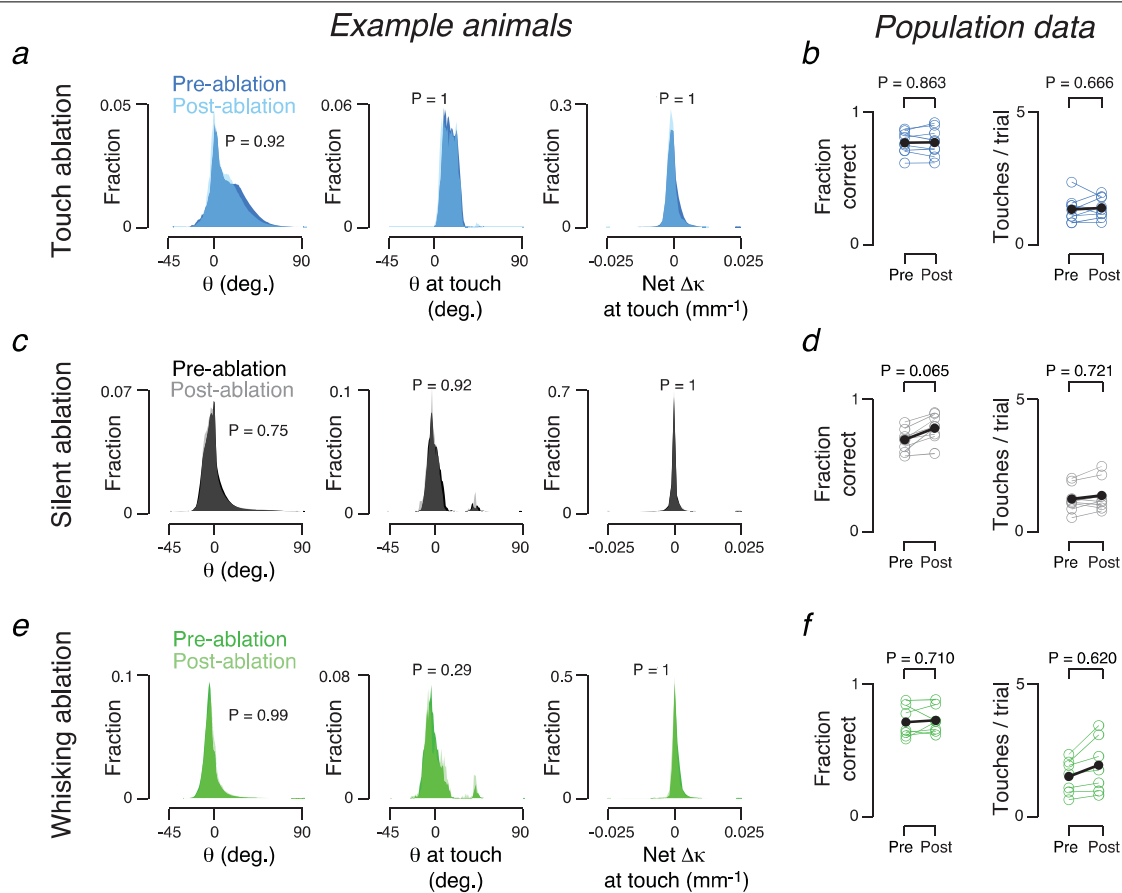
(dashed white line) in the glial antibody image along lines emanating from the ablation centre at varied angles (Methods). Top, ridge along ablation from **b**. Bottom left, intensity image in angle-distance space within which the ridge was measured. Bottom right, distribution of reaction radii; all points constitute IBA1 labelling, as no detectable glial reactions were observed with GFAP: 8 out of 10 IBA1-labelled and 0 out of 6 GFAP-labelled ablations retrieved histologically revealed a detectable glial reaction (Methods). These reactions had radii of $11.7 \pm 1.7 \mu\text{m}$ (mean \pm s.d.; $n = 8$ sites). **e**, Two-photon in vivo images before (left) and after (right) a successful ablation (target, white dotted line). Green, GCaMP6s fluorescence; red, mCherry fluorescence. **f**, As in **e**, but after a failure of the ablation protocol to terminate the ablation. Excess energy deposition produced a large lesion (black, centre of image). White arrows denote corresponding points in the two images. Mice ($n = 5$) with such lesions were excluded from the study.



Extended Data Fig. 4 | Effect of ablation on L2/3 model sensory representation increases with the number of ablated neurons. Ablation effect (change in R_{stimulus}) as a function of the degree of touch representation degradation (net R_{stimulus} across ablated neurons). In all cases, we used increased subnetwork connectivity (0.4) (Fig. 1, Methods). The n_{ablated} neurons with the highest encoding score were selected for ablation. Grey circles denote individual networks. Black dots denote, median across $n = 30$ simulated networks for a given number of ablated neurons, indicated in the plot. Beyond $n_{\text{ablated}} > 50$, we observed instability, presumably because we did not attempt to restore excitatory-inhibitory balance after ablation; these data were omitted. Correlation of net R_{stimulus} ablated and $\Delta R_{\text{stimulus}}$: $R = -0.65$, $P < 0.01$.



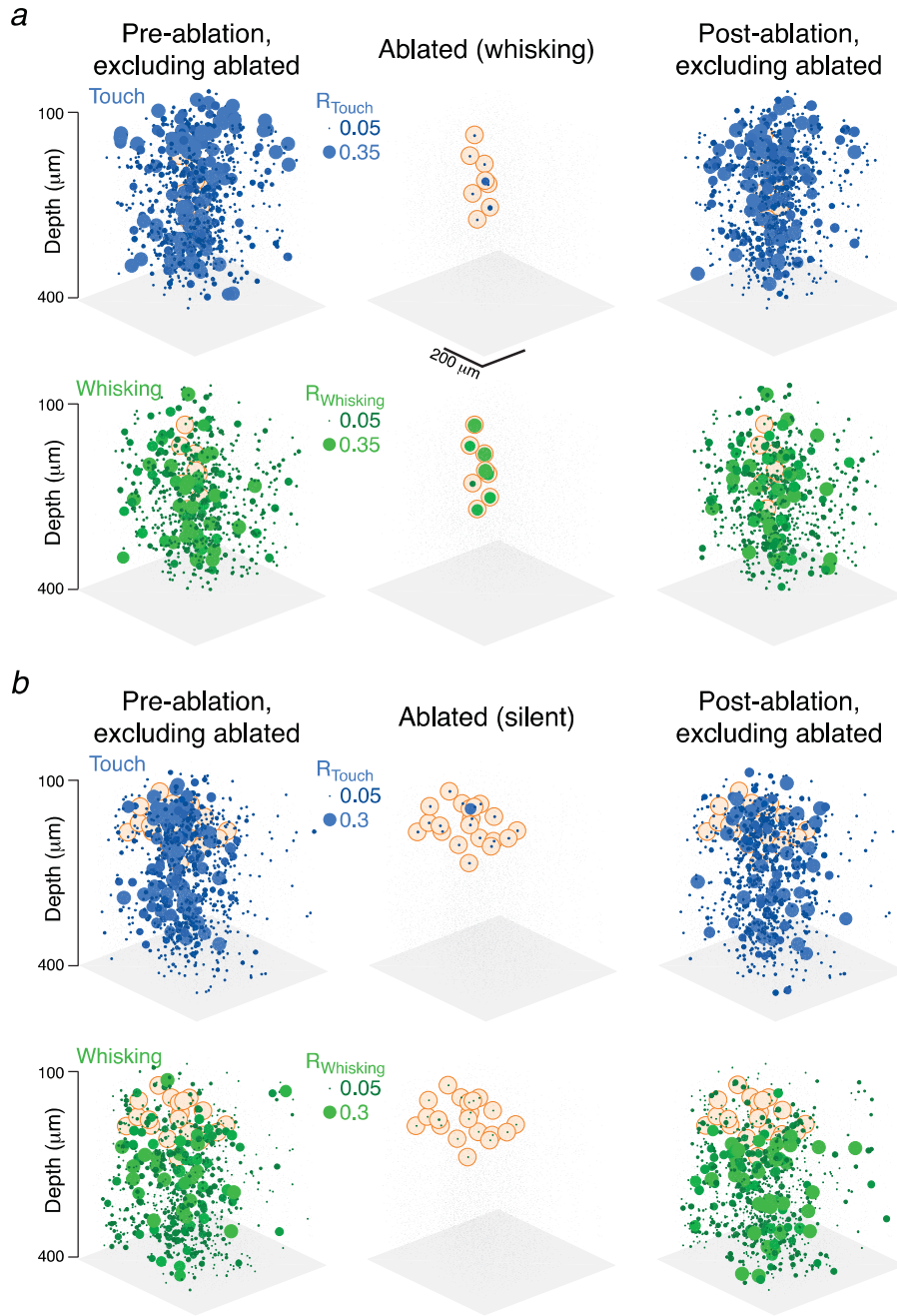
Extended Data Fig. 5 | Ablation of touch neurons does not produce distance-dependent effects in the whisking representation. Proximal (15–35 μm to nearest ablated cell) change in P_{whisking} : 0.000 ± 0.192 (grand median \pm adjusted MAD); distal (115–135 μm): -0.023 ± 0.186 . P value given for Wilcoxon signed-rank test comparing proximal paired to distal ($n=9$ mice). Legend as in Fig. 2m.



Extended Data Fig. 6 | Behaviour does not account for ablation effects.

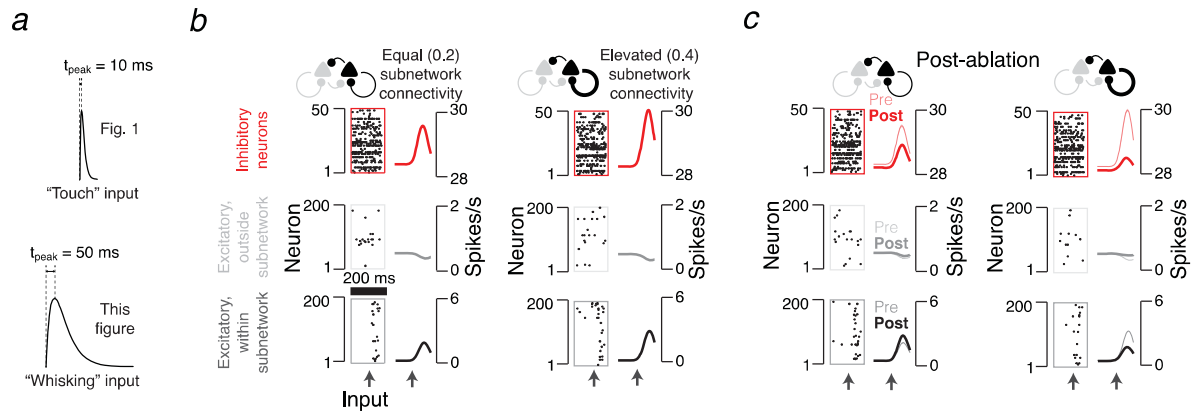
a, Pre-touch (dark blue) and post-touch (light blue) distributions of neuron ablation for whisker angle (θ), angle at touch (θ at touch), and net curvature change across all touches (net $\Delta\kappa$ at touch) in an example mouse. P values from a Kolmogorov-Smirnov test comparing the distribution of a variable before and after ablation in individual mice. No mice showed a significant ($P < 0.05$)

change in any of the three parameters. **b**, Fraction of correct trials (left) and number of touches (right) before and after ablation of touch neurons. P values from a Wilcoxon signed-rank test, paired by mouse ($n = 9$ mice). **c, d**, As in **a** and **b**, but for ablations of silent neurons ($n = 8$ mice). **e, f**, As in **a** and **b**, but for ablations of whisking neurons ($n = 7$ mice).



Extended Data Fig. 7 | Example effects of ablations of whisking and silent neurons. a, Example ablation of a whisking neuron. Left, example maps for touch (top, blue) and whisking (bottom, green) cells before ablation. Sphere size corresponds to R_{touch} (top) or R_{whisking} (bottom) values. Grey dots denote other neurons. These maps exclude the ablated neurons, the position of which

is indicated by a faint orange background. Centre, R_{touch} (top) and R_{whisking} (bottom) values for the ablated population. Right, R_{touch} (top) and R_{whisking} (bottom) values after ablation, with ablated neurons again excluded. **b**, As in **a**, but for ablation of silent neurons.



Extended Data Fig. 8 | Simulation of ablation of whisking neurons produces representation degradation in networks with increased, but not equal subnetwork connectivity. **a**, Whisking input was simulated by using input with a peak time of 50 ms (bottom), in contrast to 10 ms for touch (top) (Fig. 1). **b**, Model network responses aligned to input before ablation. Left to right, increasing subnetwork connectivity. Bottom to top, raster plots (each showing

a subset of neurons from an example network) and PSTHs (averaged across all neurons and networks) for the three neuronal populations in the model (Fig. 1a, Methods). **c**, As in **b**, but after ablation. Thin PSTHs are before ablation. Change in encoding score after ablation: from 0.143 ± 0.024 to 0.104 ± 0.013 , grand median \pm adjusted MAD. $P < 0.001$, Wilcoxon signed-rank test paired by network ($n = 30$ networks).

Extended Data Table 1 | Individual mice are listed with the number of neurons imaged, the type(s) of ablation(s) performed, and the number of neurons ablated

Animal ID	Cell count	Ablated type	Ablated count	Touch count	ΔR_{touch}	Whisking count	$\Delta R_{\text{whisking}}$
j250220	8,844	touch	50	867	-0.122	422	-0.076
j257218	6,215	silent	20	743	0.018	366	0.032
j257220	3,663	touch	9	389	-0.004	455	0.025
j258836	7,290	silent	16	481	0.066	666	-0.012
		touch	15	488	-0.055	655	0.001
j271211	9,415	silent	16	2,476	-0.012	551	-0.007
		touch	15	1,951	-0.061	692	0.021
j278937	4,230	silent	15	146	0.004	379	0.001
		whisking	12	150	-0.006	392	0.011
j278939	7,167	silent	20	636	-0.024	698	0.001
		whisking	13	523	-0.017	649	-0.015
j281915	7,002	whisking	8	723	-0.010	550	0.004
		touch	11	601	-0.022	672	0.008
n274424	7,785	touch	17	706	-0.029	718	-0.014
n272761	6,359	touch	14	509	-0.001	410	0.000
n274577	12,173	touch	11	1,380	-0.013	657	0.004
n275801	8,354	silent	14	220	0.000	738	-0.007
		touch	9	197	-0.065	698	-0.006
n275798	11,364	silent	13	959	0.002	958	0.002
		whisking	11	972	0.008	1,037	0.011
n278288	9,314	whisking	25	969	-0.008	813	-0.006
n276013	11,659	silent	16	652	0.028	1,072	-0.001
		whisking	9	634	-0.012	968	-0.008
n278759	9,187	whisking	11	617	-0.003	526	-0.004

The number of touch and whisking neurons is given (for the 0.1 encoding score criteria) (Methods), as well as the change in median encoding score after ablation. Mice with IDs starting with a 'j' were virally transfected *Emx1-cre* \times *LSL-H2B-mCherry*⁷; mice with IDs starting with an 'n' were *Ai162* \times *Slc17a7-IRES-cre*⁴⁰.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Scanimage was used to collect imaging data. Norpix, along with custom MATLAB code was used for videography data collection. Behavior control and data collection was via BPod and custom MATLAB code.

Data analysis

Data analysis was performed with custom MATLAB code.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data are available at crcns.org.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples sizes were selected based on typical sample sizes in the field. 7-9 mice were used per ablation type. No statistical tests were used to determine sample size.
Data exclusions	Animals with excessive damage were excluded. Furthermore, neurons immediately adjacent to the ablated cells were excluded from analysis. These criteria were determined prior to experiments, and are described in detail in the manuscript.
Replication	We did 7-9 replications of each experiment. All presented results were reproducible across replicates.
Randomization	Mice were assigned randomly to experimental groups (ablation type).
Blinding	Experimenters were not blind to the group each mouse belonged to.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	We used antibodies for microglial marker iba1 (Abcam, ab5076) or the astrocytic marker GFAP (Abcam, ab7260).
Validation	Validation was implicit. Microglia and astrocytes typically have uniform reactivity in cortex. We looked for, and observed, this pattern in tissue that was not subject to ablation.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Transgenic mice are listed in the table, Extended Data Table 1. We used two types of mice: a cross between R26-LSL-H2B-mCherry (JAX 023139) and exm1-Cre (JAX 005628), as well as crosses of Ai162 (JAX 031562) and Slc17a7-IRES2-Cre (JAX 023527).
Wild animals	N/A
Field-collected samples	N/A
Ethics oversight	All procedures were performed in compliance with the Janelia Research Campus Institutional Animal Care and Use Committee and the New York University University Animal Welfare Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Decoy exosomes provide protection against bacterial toxins

<https://doi.org/10.1038/s41586-020-2066-6>

Received: 7 December 2018

Accepted: 9 January 2020

Published online: 4 March 2020



Matthew D. Keller^{1,2}, Krystal L. Ching^{1,2}, Feng-Xia Liang^{3,4}, Avantika Dhabaria^{3,5}, Kayan Tam¹, Beatrix M. Ueberheide^{3,5,6}, Derya Unutmaz⁷, Victor J. Torres^{1,9}✉ & Ken Cadwell^{1,2,8,9}✉

The production of pore-forming toxins that disrupt the plasma membrane of host cells is a common virulence strategy for bacterial pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA)^{1–3}. It is unclear, however, whether host species possess innate immune mechanisms that can neutralize pore-forming toxins during infection. We previously showed that the autophagy protein ATG16L1 is necessary for protection against MRSA strains encoding α -toxin⁴—a pore-forming toxin that binds the metalloprotease ADAM10 on the surface of a broad range of target cells and tissues^{2,5,6}. Autophagy typically involves the targeting of cytosolic material to the lysosome for degradation. Here we demonstrate that ATG16L1 and other ATG proteins mediate protection against α -toxin through the release of ADAM10 on exosomes—extracellular vesicles of endosomal origin. Bacterial DNA and CpG DNA induce the secretion of ADAM10-bearing exosomes from human cells as well as in mice. Transferred exosomes protect host cells in vitro by serving as scavengers that can bind multiple toxins, and improve the survival of mice infected with MRSA in vivo. These findings indicate that ATG proteins mediate a previously unknown form of defence in response to infection, facilitating the release of exosomes that serve as decoys for bacterially produced toxins.

We previously demonstrated that primary cells obtained from mice with hypomorphic expression of *Atg16l1* (*Atg16l1tm*) display an increase in total ADAM10 levels and are susceptible to lysis when cultured in the presence of α -toxin⁴. Consistent with these findings, we have now found that levels of cell-surface and total ADAM10 are increased in the human alveolar epithelial cell line A549 upon short hairpin (sh)RNA-mediated depletion of *ATG16L1* (Fig. 1a–d). *ATG16L1*-knockdown cells treated with purified α -toxin displayed increased cell death compared with control cells transduced with nontargeting shRNA, whereas *ADAM10*-knockdown cells were resistant (Fig. 1e). *ATG16L1* mediates the conjugation of phosphatidylethanolamine to the ubiquitin-like molecule LC3—a step that is necessary for the proper biogenesis of the autophagosome and for subsequent events in which substrates are degraded by the lysosome⁷. Inhibiting ULK1, a kinase upstream of *ATG16L1*, or *ATG5*, a binding partner of *ATG16L1*, led to increased cell-surface ADAM10 levels similar to those produced by knocking down *ATG16L1* (Fig. 1f). Prevention of lysosomal acidification by weak bases alters endosomal recycling to the plasma membrane^{8,9}. Although total levels of ADAM10 and the autophagy substrate SQSTM1 were increased when A549 cells were treated with lysosomal acidification inhibitors (NH₄Cl, chloroquine or bafilomycin), all three agents decreased surface ADAM10 levels (Fig. 1g, h and Extended Data Fig. 1a–d). Surface levels of epithelial cell adhesion molecule (EpcAM) were unaltered, indicating that lysosome inhibition did not affect all plasma-membrane

molecules (Extended Data Fig. 1e–g). ADAM10 levels were unaffected by proteasome inhibition (Extended Data Fig. 1h, i), suggesting that ATG proteins reduce cell-surface ADAM10 through a lysosome- and proteasome-independent process.

ATG proteins mediate the extracellular release of soluble and vesicle-bound substrates through a process broadly referred to as secretory autophagy¹⁰. ADAM10 is known to be incorporated into exosomes—extracellular vesicles typically 40–120 nm in diameter^{11,12}. Thus, we hypothesized that the autophagy machinery prevents ADAM10 accumulation on cells by facilitating its secretion on exosomes. We found a reduction in the lower-molecular-weight band of ADAM10 (a mature form cleaved during trafficking from the endoplasmic reticulum) in exosome fractions isolated from the culture supernatants of *ATG16L1*-knockdown cells compared with control cells treated with nontargeting shRNA (Fig. 2a, b and Extended Data Fig. 2a). Western blot analysis confirmed that the fractionation procedure led to enrichment of the exosomal marker CD9 and not the microvesicle marker ARF6 (Extended Data Fig. 2b). Parallel analysis by transmission electron microscopy (TEM) indicated that the exosome fraction contained a greater number of single-lipid-bilayer vesicles of 80–150 nm in diameter compared with microvesicles larger than 150 nm (Extended Data Fig. 2c–e).

The decrease in ADAM10 that occurs after *ATG16L1* inhibition reflects a general reduction in exosome levels: we observed a reduction in CD9 levels by western blot and a reduction in the number of vesicles by TEM

¹Department of Microbiology, New York University School of Medicine, New York, NY, USA. ²Kimmel Center for Biology and Medicine at the Skirball Institute, New York University School of Medicine, New York, NY, USA. ³Division of Advanced Research Technologies, New York University Langone Health, New York, NY, USA. ⁴The Microscopy Laboratory at New York University Langone Health, New York, NY, USA. ⁵The Proteomics Laboratory at New York University Langone Health, New York, NY, USA. ⁶The Laura and Isaac Perlmutter Cancer Center, New York, NY, USA. ⁷Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. ⁸Division of Gastroenterology and Hepatology, Department of Medicine, New York University Langone Health, New York, NY, USA. ⁹These authors contributed equally: Victor J. Torres, Ken Cadwell. ✉e-mail: Victor.Torres@nyulangone.org; Ken.Cadwell@nyulangone.org

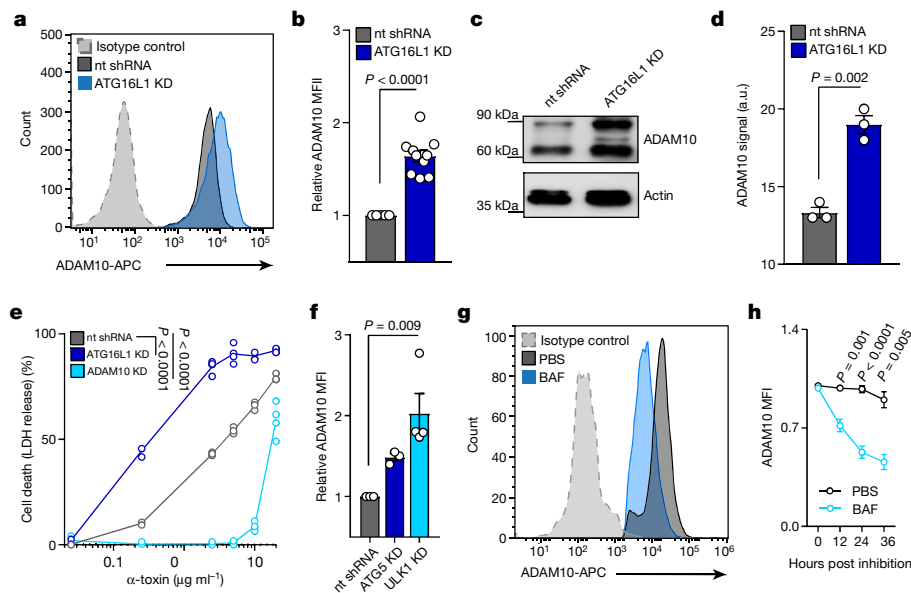


Fig. 1 | ATG16L1 inhibits surface ADAM10 independently of lysosomal degradation. **a, b**, Representative flow-cytometry histogram (**a**) and quantification of mean fluorescent intensity (MFI) (**b**) of surface ADAM10 in A549 cells following *ATG16L1* knockdown (ATG16L1 KD; $n = 5$); or in cells containing nontargeting control shRNA (nt shRNA; $n = 10$). **c, d**, Representative western blot (**c**) and quantification (**d**) of ADAM10 in ATG16L1 KD and control cells; $n = 3$. a.u., arbitrary units. **e**, Quantification of cell death (assayed by release of lactate dehydrogenase, LDH) of nt shRNA, ATG16L1 KD and ADAM10 KD cells following treatment with purified α -toxin; $n = 4$. **f**, Quantification of

surface ADAM10 by flow cytometry in nt shRNA ($n = 3$), ATG5 KD ($n = 3$) and ULK1 KD ($n = 4$) A549 cells. **g, h**, Representative flow-cytometry histogram from three independent repeats of surface ADAM10 on A549 cells 24 h after treatment with bafilomycin (BAF; 10 nM) (**g**), and quantification of MFI over time following addition of BAF (**h**; $n = 3$). PBS, phosphate-buffered saline. Measurements were taken from distinct samples, and graphs show means and standard errors of the mean (s.e.m.). **b, d, f, h**, Two-tailed, unpaired *t*-test with Welch's correction compared with nt shRNA controls. **e**, Two-tailed, unpaired *t*-test of area under curve compared with nt shRNA controls.

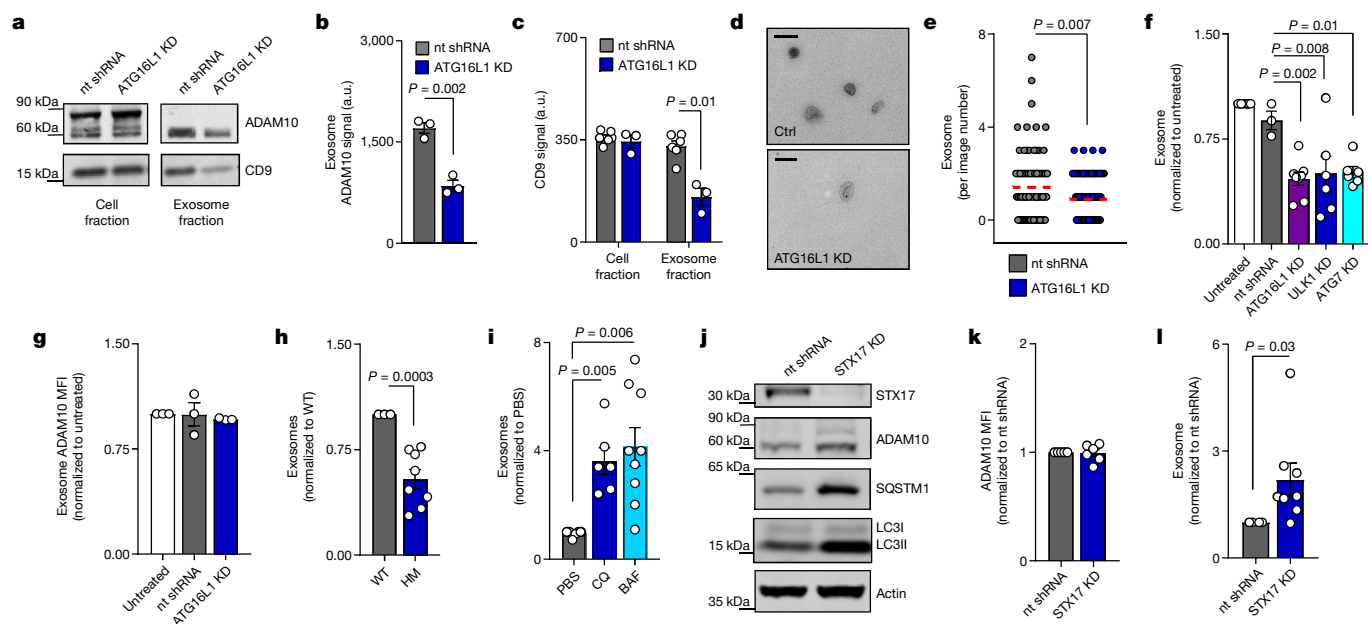


Fig. 2 | ATG proteins regulate the release of ADAM10-containing exosomes. **a–c**, Representative ADAM10 and CD9 western blot from three independently repeated experiments (**a**); quantification of exosome ADAM10, $n = 3$ (**b**); and quantification of CD9 in cell lysates and exosomes from nt shRNA ($n = 7$) and ATG16L1 KD ($n = 3$) cells (**c**). **d, e**, Representative transmission electron micrographs (**d**) and quantification (**e**) of vesicles in the exosome fraction of nt shRNA and ATG16L1 KD culture supernatants. Scale bars, 100 μ m; $n = 80$ images. Ctrl, control. **f**, Flow-cytometric quantification of exosomes from untreated ($n = 4$), nt shRNA ($n = 3$), ATG16L1 KD ($n = 7$), ULK1 KD ($n = 6$) and ATG7 KD ($n = 6$) A549 cells. **g**, Quantification of ADAM10 MFI in untreated, nt shRNA and ATG16L1 KD exosomes from **f**. $n = 3$. **h**, Exosome quantification (CD9⁺, CD63⁺, CD81⁺ and PKH67⁺ structures) in blood from C57BL/6J (wild-type (WT);

$n = 6$) and ATG16L1 hypomorph (HM; $n = 8$) mice. **i**, Exosome quantification following addition of PBS ($n = 4$), chloroquine (CQ; $n = 5$) or BAF ($n = 9$). **j**, Representative western blot from three independent repeats analysing ADAM10, SQSTM1 and LC3II levels in nt shRNA and STX17 KD cells. **k**, ADAM10 MFI of nt shRNA ($n = 5$) and STX17 KD ($n = 8$) cells. **l**, Exosome quantification from nt shRNA and STX17 KD cells; $n = 8$. Measurements were taken from distinct samples and graphs show means \pm s.e.m. **b, c, e, h, i**, Two-tailed, unpaired *t*-test with Welch's correction compared with nt shRNA or WT controls. **f, g, i**, One-way analysis of variance (ANOVA) with Dunnett's post-test compared with nt shRNA or PBS. Data represent at least three independent experiments.

in the exosome fraction of *ATG16L1*-knockdown cell-culture supernatant (Fig. 2a, c–e). To further validate these results through a quantitative assay, we used flow cytometry in which antibody-based staining of the surface exosome markers CD9, CD63 and CD81 was combined with PKH67, a fluorescent lipid-bilayer-intercalating compound (Extended Data Fig. 2f). Depletion of autophagy proteins substantially reduced the total numbers of exosomes in the culture supernatant (Fig. 2f). *ATG16L1*-knockdown reduced the total number of ADAM10-positive exosomes but not the amount of ADAM10 per exosome (Fig. 2g), confirming that the ATG proteins regulate exosome biogenesis rather than substrate incorporation. We also found that the blood from *Atg16l1*^{HM} mice contained fewer exosomes than blood from wild-type mice (Fig. 2h).

Our finding that blocking lysosomal acidification decreases plasma-membrane ADAM10 levels could be explained by a mechanism in which inhibiting late-stage autophagy redirects the autophagy machinery towards generation of exosomes^{13–15}. Consistent with this possibility, we detected increases in CD9 and ADAM10 levels in the exosome fraction as well as an increase in total exosome numbers in the culture supernatant of cells treated with chloroquine or bafilomycin (Fig. 2i and Extended Data Fig. 3a–d). The SNARE protein syntaxin 17 (STX17) mediates autophagosome–lysosome fusion and is dispensable for secretory autophagy^{16,17}. *STX17* knockdown increased total ADAM10, SQSTM1 and LC3II levels, indicating successful inhibition of autophagy, without increasing surface levels of ADAM10 (Fig. 2j, k). However, supernatants from *STX17*-knockdown cells contained more exosomes (Fig. 2l), indicating that ATG proteins mediate the release of exosomes in a manner distinct from conventional degradative autophagy.

We next examined whether ATG-dependent exosome production is induced by pathogen exposure. Heat-killed *S. aureus* (CA-MRSA USA300, hereafter HKSA), an isogenic α -toxin-deficient USA300 strain (Δhla), *Streptococcus pneumoniae*, *Citrobacter rodentium* and *Salmonella enterica* Typhimurium all increased exosome production in human and mouse cells (Fig. 3a and Extended Data Fig. 4a, i). After testing several bacterially derived products, we identified bacterial DNA and CpG DNA as the exosome inducer (Fig. 3b and Extended Data Fig. 4b–g). Furthermore, addition of DNA isolated from *S. aureus* to cells elicited exosomes, and DNase treatment abolished this effect (Extended Data Fig. 4j). Exosome production in response to HKSA and CpG DNA depended on the endosomal DNA-sensor Toll-like receptor 9 (TLR9) (Fig. 3c and Extended Data Fig. 4h). Inducing autophagy with Torin-1—an inhibitor of mammalian target of rapamycin (mTOR)—did not induce exosomes, suggesting that TLR9 acts through a distinct mechanism (Extended Data Fig. 4k, l). Instead, the addition of CpG DNA or bafilomycin (a positive control) individually or together decreased LysoSensor staining, an indicator of acidic organelles (Extended Data Fig. 5a–c). We also found that treating cells with the neutral sphingomyelinase inhibitor GW4869—which prevents the generation of vesicles that become exosomes by interfering with the inward budding of the multivesicular body (MVB)¹⁸—impairs CpG-DNA-induced exosome production (Extended Data Fig. 4m). Thus, the membrane-trafficking events downstream of TLR9 probably contribute to exosome production by regulating endosomal trafficking and vesicle-biogenesis events that include the MVB.

Intravenous injection of heat-killed or live *S. aureus* into wild-type mice led to a marked increase in the number of exosomes in their blood that was blunted in *Atg16l1*^{HM} mice, but not in mice in which *Atg16l1* was selectively deleted in macrophage and dendritic-cell lineages (Fig. 3d, e and Extended Data Fig. 4n, o). This observation is consistent with our previous study in which *Atg16l1*^{HM} mice, but not myeloid-cell-specific *Atg16l1* knockout mice, were susceptible to lethal bloodstream infection by MRSA⁴. Next, we performed mass spectrometry on exosomes from the blood of mice inoculated intranasally with HKSA or CpG DNA (Extended Data Fig. 4p). The majority of detected proteins originated from the liver and were previously identified in exosomes and extracellular spaces (Fig. 3f–h and Supplementary Tables 1–3). We confirmed that the liver

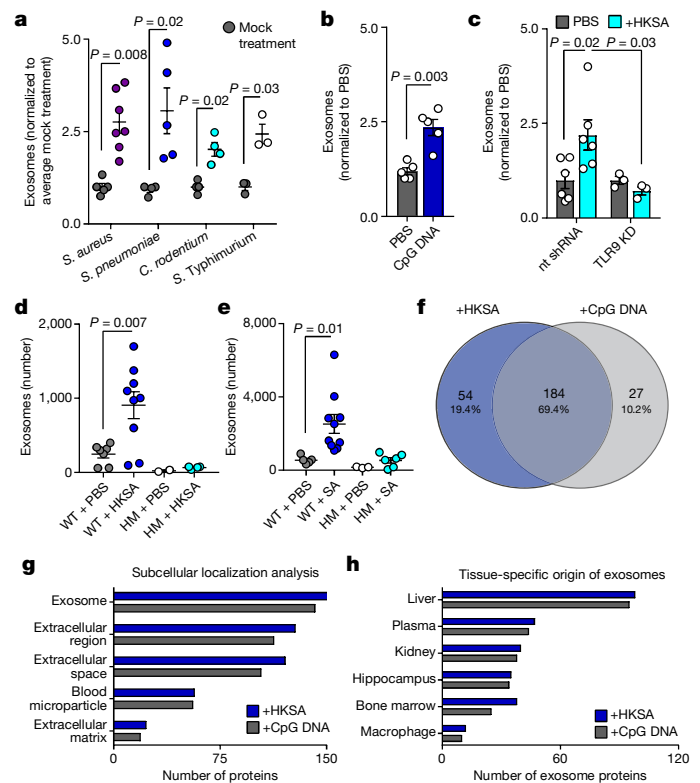


Fig. 3 | Bacteria induce exosome production. **a–e**, Flow-cytometric quantification of exosomes in A549 cell-culture supernatant 18 h after exposure to heat-killed *S. aureus* (n = 7), heat-killed *S. pneumoniae* (n = 5), heat-killed *C. rodentium* (n = 4), heat-killed *S. Typhimurium* (n = 3) (a); after CpG DNA treatment (4 μ M; n = 5) (b); in nt shRNA (n = 6) and TLR9 shRNA (TLR9 KD; n = 3) targeted A549 cells following HKSA exposure (c); in blood from wild-type and *Atg16l1*^{HM} mice following intranasal (i.n.) inoculation with HKSA (1×10^8 colony-forming units (CFU); WT plus PBS, n = 7; WT plus HKSA, n = 9; HM plus PBS, n = 2; HM plus HKSA, n = 4) (d); or following intravenous (i.v.) inoculation with live *S. aureus* (1×10^7 CFU; WT plus PBS, n = 5; WT plus HKSA, n = 10; HM plus PBS, n = 3; HM plus HKSA, n = 6) (e). **f**, Venn diagram of shared and discrete proteins identified by mass spectrometry in exosomes isolated from the blood of mice exposed to HKSA or CpG DNA i.n. (1×10^8 CFU; 20 μ g CpG DNA). **g**, Gene-ontology analysis of the subcellular location of proteins identified by mass spectrometry. **h**, Tissue-specific origin of exosome proteins. Measurements were taken from distinct samples and graphs show means \pm s.e.m. **a**, **b**, Two-tailed, unpaired *t*-test with Welch's correction compared with PBS controls. **c**–**e**, One-way ANOVA with Dunnett's post-test compared with nt shRNA plus PBS, or WT plus PBS controls.

enzyme argininosuccinate synthase 1 (ASS1) was enriched in HKSA, and that CpG DNA elicited exosomes *in vivo*¹⁹ (Extended Data Fig. 4p).

Next, we tested whether these released vesicles could serve as a host response to bind and inhibit toxins. We found that exosomes isolated from control donor cells, but not from *ATG16L1*-knockdown cells, were able to protect A549 target cells from α -toxin toxicity (Fig. 4a). Adding twice the volume of the supernatant of *ATG16L1*-knockdown cells from which exosomes were isolated improved the viability of the cells, indicating that the inability of exosomes from *ATG16L1*-knockdown cells to protect cells was due to a reduction in the number of exosomes. Exosomes harvested from *ADAM10*-knockdown cells were unable to protect cells (Fig. 4a and Extended Data Fig. 6a). These results were confirmed with exosomes purified through fluorescence-activated cell sorting (FACS; Fig. 4b). Of note, preincubating cells with HKSA or CpG DNA also protected against α -toxin toxicity (Extended Data Fig. 6b). This protection was due to exosomes: removing the exosome-containing supernatant restored susceptibility to α -toxin in HKSA

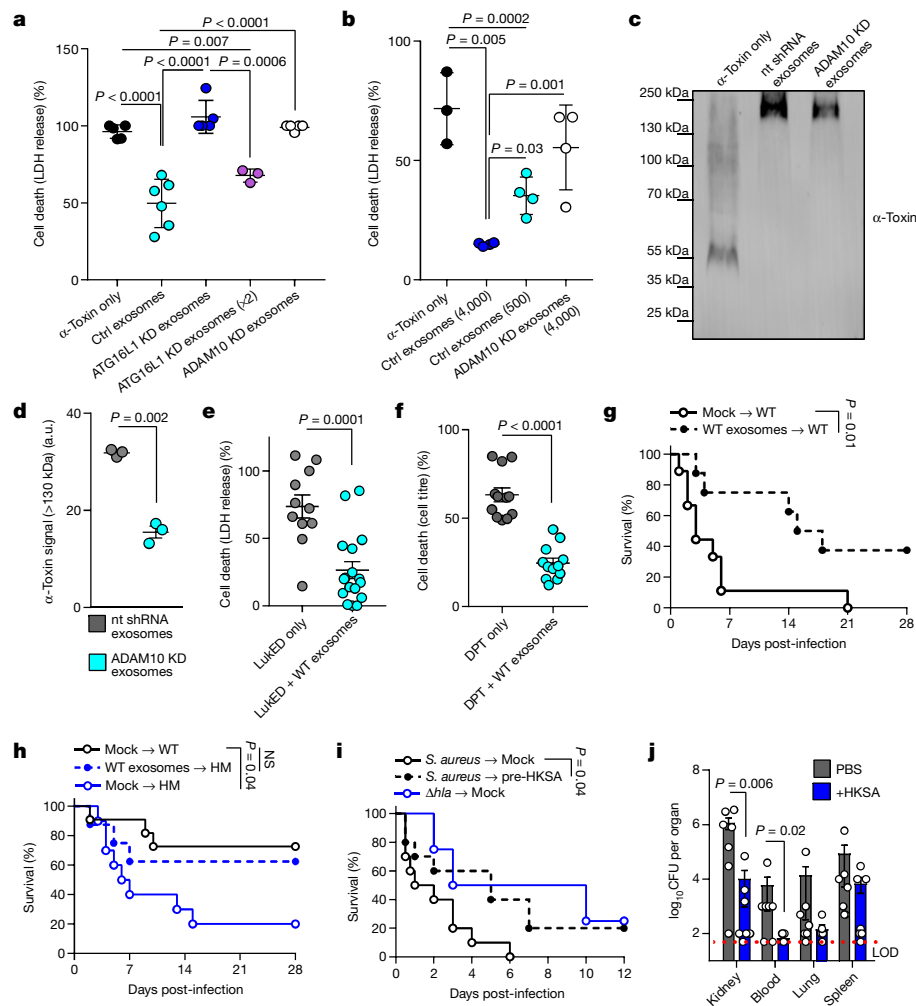


Fig. 4 | Exosomes protect against bacterial toxins. a, b, A549 cell death following treatment with α -toxin together with exosomes isolated from nt shRNA ($n=6$), ATG16L1 KD ($n=6$), ATG16L1 KD \times 2 ($n=3$), ADAM10 KD ($n=5$) cells (**a**); or with FACS-purified exosomes ($n=4$) (**b**). **c, d**, Representative western blot (**c**) and quantification (**d**) of oligomerized α -toxin larger than 130 kDa following addition of exosomes isolated from WT or ADAM10 KD cells. $n=3$. **e**, BMDM death following treatment with LukED and exosomes isolated from WT BMDM cultures (LukED only, $n=10$; LukED plus WT exosomes, $n=16$). **f**, A549 cell death following exposure to diphtheria toxin (DPT) and exosomes isolated from A549 cultures. $n=12$. **g**, Survival of WT mice infected i.v. with *S. aureus* (USA300; 5×10^7 CFU) mock-treated or injected intraperitoneally with exosomes from WT mice. $n=9$ mice per condition. **h**, Survival of WT ($n=10$) and

Atg16l1^{HM} (Mock to HM, $n=10$; WT exosomes to HM, $n=10$) mice infected i.v. with 2.5×10^7 CFU of *S. aureus* and receiving exosomes from WT mice. NS, not significant. **i**, Survival of WT mice ($n=10$) pretreated with intranasal HKSA followed by a lethal dose of *S. aureus* (strain USA300; 5×10^7 CFU; $n=10$) or an isogenic α -toxin-deficient strain (Δhla ; $n=5$). **j**, *S. aureus* burden 24 h after infection with 1×10^7 CFU of USA300 i.v. in kidney, spleen, lung and blood (per millilitre) in mice pretreated with PBS or HKSA i.v.; $n=6$. Measurements were taken from distinct samples. LOD, limit of detection. Graphs show means \pm s.e.m. **a, b**, One-way ANOVA with Dunnett's post-test compared with α -toxin only or control exosomes. **d–f, j**, Two-tailed, unpaired *t*-test with Welch's correction compared with nt shRNA exosomes, α -toxin only or PBS controls. **g–i**, log-rank Mantel–Cox test.

or CpG-DNA-treated cells (Extended Data Fig. 6b). Exosomes seem to protect cells by inducing toxin oligomerization on the exosome membranes (Fig. 4c, d).

Exosomes elicited from mouse bone-marrow-derived macrophages (BMDMs) harbour the toxin receptor CCR5 and protected BMDMs from LukED (Fig. 4e and Extended Data Fig. 6c), another toxin produced by *S. aureus*²⁰. Similarly, exosomes isolated from A549 cells protected target cells from diphtheria toxin (Fig. 4f), a potent toxin produced by *Corynebacterium diphtheriae* that binds to the epidermal growth factor receptor (EGFR)²¹, which was present in our exosome proteomics dataset (Supplementary Table 1). Thus, exosomes can neutralize different types of toxin.

To test whether exosomes are protective in vivo, we injected donor mice with HKSA to elicit exosomes in the blood; we then transferred these exosomes into recipient mice and infected the animals intravenously with a lethal dose of *S. aureus*. Transfer of exosomes from

wild-type but not *Atg16l1*^{HM} donors extended the survival of *S. aureus*-infected wild-type recipient mice (Fig. 4g and Extended Data Fig. 6d, e). Furthermore, transfer of exosomes from a wild-type donor improved the survival of *Atg16l1*^{HM} mice injected with lethal dose 50 of *S. aureus* to levels similar to those of mock-treated wild-type mice (Fig. 4h), suggesting that the increased susceptibility of *Atg16l1* mutants is in part due to reduced exosome production.

Finally, priming mice with intravenous injection of HKSA increased the relative amount of α -toxin oligomers to monomers in the exosome fraction isolated from the bronchoalveolar lavage (BAL) fluid compared with mock-treated controls (Extended Data Fig. 6f–k). Additionally, we found that conditioning the mice with HKSA prolonged survival following *S. aureus* infection, phenocopying control mice infected with the Δhla strain (Fig. 4i). To monitor bacterial burden, we challenged mice with a lower inoculum, and found that conditioning with HKSA resulted in reduction in *S. aureus* burdens in the kidneys and blood (Fig. 4j).

Exosomes extend the functional reach of their cells of origin and contribute to a spectrum of biological processes^{22,23}. Our results suggest that exosomes can also provide a previously undescribed innate immune response to bacterial infection by acting as decoys that trap membrane-acting virulence factors, such as pore-forming toxins, to prevent injury of target tissues. The action of these ‘defensosomes’ is in line with recent evidence that engineered liposomes can neutralize toxins from Gram-positive bacteria²⁴. Our findings also indicate that ATG proteins regulate the production of exosomes during host defence—a mechanism that is distinct from the recently uncovered roles of ATG16L1 in promoting plasma-membrane repair during *Listeria monocytogenes* infection, or in preventing necroptosis mediated by the pore-forming molecule MLKL during norovirus infection^{25,26}. Rather, our findings resemble recently described processes in which extracellular vesicles are produced when the atypical ATG12–ATG3 conjugate promotes MVB function, or when ATG5 disrupts lysosomal acidification^{27,28}. Given that the origin and regulation of extracellular vesicles remain poorly defined, a detailed understanding of the cellular response triggering the production of defensosomes during infection may reveal opportunities to leverage their unique properties to combat bacteria and other pathogens.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2066-6>.

1. Dal Peraro, M. & van der Goot, F. G. Pore-forming toxins: ancient, but never really out of fashion. *Nat. Rev. Microbiol.* **14**, 77–92 (2016).
2. Seillie, E. S. & Bubeck Wardenburg, J. *Staphylococcus aureus* pore-forming toxins: the interface of pathogen and host complexity. *Semin. Cell Dev. Biol.* **72**, 101–116 (2017).
3. Spaan, A. N., van Strijp, J. A. G. & Torres, V. J. Leukocidins: staphylococcal bi-component pore-forming toxins find their receptors. *Nat. Rev. Microbiol.* **15**, 435–447 (2017).
4. Maurer, K. et al. Autophagy mediates tolerance to *Staphylococcus aureus* alpha-toxin. *Cell Host Microbe* **17**, 429–440 (2015).
5. Inoshima, I. et al. A *Staphylococcus aureus* pore-forming toxin subverts the activity of ADAM10 to cause lethal infection in mice. *Nat. Med.* **17**, 1310–1314 (2011).
6. Becker, R. E., Berube, B. J., Sampedro, G. R., DeDent, A. C. & Bubeck Wardenburg, J. Tissue-specific patterning of host innate immune responses by *Staphylococcus aureus* α-toxin. *J. Innate Immun.* **6**, 619–631 (2014).
7. Matsuzawa-Ishimoto, Y., Hwang, S. & Cadwell, K. Autophagy and inflammation. *Annu. Rev. Immunol.* **36**, 73–101 (2018).

8. Presley, J. F., Mayor, S., McGraw, T. E., Dunn, K. W. & Maxfield, F. R. Bafilomycin A1 treatment retards transferrin receptor recycling more than bulk membrane recycling. *272*, 13929–13936 (1997).
9. Johnson, L. S., Dunn, K. W., Pytowski, B. & McGraw, T. E. Endosome acidification and receptor trafficking: bafilomycin A1 slows receptor externalization by a mechanism involving the receptor’s internalization motif. **4**, 1251–1266 (1993).
10. Cadwell, K. & Debnath, J. Beyond self-eating: the control of nonautophagic functions and signaling pathways by autophagy-related proteins. *J. Cell Biol.* **217**, 813–822 (2018).
11. Stoeck, A. et al. A role for exosomes in the constitutive and stimulus-induced ectodomain cleavage of L1 and CD44. *Biochem. J.* **393**, 609–618 (2006).
12. Shimoda, M. et al. Loss of the Timp gene family is sufficient for the acquisition of the CAF-like cell state. *Nat. Cell Biol.* **16**, 889–901 (2014).
13. Ejlerskov, P. et al. Tubulin polymerization-promoting protein (TPPP/p25a) promotes unconventional secretion of α-synuclein through exophagy by impairing autophagosome-lysosome fusion. *J. Biol. Chem.* **288**, 17313–17335 (2013).
14. Guo, H. et al. Atg5 disassociates the V1V0-ATPase to promote exosome production and tumor metastasis independent of canonical macroautophagy. *Dev. Cell* **43**, 716–730 (2017).
15. Miao, Y., Li, G., Zhang, X., Xu, H. & Abraham, S. N. A TRP channel senses lysosome neutralization by pathogens to trigger their expulsion. *Cell* **161**, 1306–1319 (2015).
16. Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151**, 1256–1269 (2012).
17. Kimura, T. et al. Dedicated SNAREs and specialized TRIM cargo receptors mediate secretory autophagy. *EMBO J.* **36**, 42–60 (2017).
18. Trajkovic, K. et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244–1247 (2008).
19. Cho, Y. E. et al. Increased liver-specific proteins in circulating extracellular vesicles as potential biomarkers for drug- and alcohol-induced liver injury. *PLoS One* **12**, e0172463 (2017).
20. Alonzo, F., III et al. CCR5 is a receptor for *Staphylococcus aureus* leukotoxin ED. *Nature* **493**, 51–55 (2013).
21. Mitamura, T., Higashiyama, S., Taniguchi, N., Klagsbrun, M. & Mekada, E. Diphtheria toxin binds to the epidermal growth factor (EGF)-like domain of human heparin-binding EGF-like growth factor/diphtheria toxin receptor and inhibits specifically its mitogenic activity. *J. Biol. Chem.* **270**, 1015–1019 (1995).
22. Montecalvo, A. et al. Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. *Blood* **119**, 756–766 (2012).
23. Kalluri, R. The biology and function of exosomes in cancer. *J. Clin. Invest.* **126**, 1208–1215 (2016).
24. Henry, B. D. et al. Engineered liposomes sequester bacterial exotoxins and protect from severe invasive infections in mice. *Nat. Biotechnol.* **33**, 81–88 (2015).
25. Tan, J. M. J. et al. An ATG16L1-dependent pathway promotes plasma membrane repair and limits *Listeria monocytogenes* cell-to-cell spread. *Nat. Microbiol.* **3**, 1472–1485 (2018).
26. Matsuzawa-Ishimoto, Y. et al. Autophagy protein ATG16L1 prevents necroptosis in the intestinal epithelium. *J. Exp. Med.* **214**, 3687–3705 (2017).
27. Murrow, L., Malhotra, R. & Debnath, J. ATG12–ATG3 interacts with Alix to promote basal autophagic flux and late endosome function. *Nat. Cell Biol.* **17**, 300–310 (2015).
28. Guo, H. et al. Atg5 disassociates the V1V0-ATPase to promote exosome production and tumor metastasis independent of canonical macroautophagy. *Dev. Cell* **43**, 716–730 (2017).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Mice

Age- and gender-matched 8–12-week-old mice on the C57BL/6J background were used. *Atg16L1^{fl/fl}* mice on the C57BL/6J background were previously described^{4,29,30}. ‘Wild-type’ refers to littermate controls generated from breeder pairs that were heterozygous for the *Atg16L1^{fl/fl}* allele for experiments that involve comparisons between genotype. For other experiments, C57BL/6J mice were purchased from the Jackson Laboratory and bred onsite. All animal studies were performed according to approved protocols and ethical guidelines established by the NYU School of Medicine Institutional Animal Care and Use Committee (IACUC) and Institutional Review Board.

Cell lines

For in vitro studies, the human lung epithelial cell line, A549, was purchased from ATCC (catalogue number CCL-185). Human embryonal kidney cells, HEK293FT, purchased from ThermoFisher (catalogue number R70007), were used for lentiviral packaging. All cell lines were confirmed as free from mycoplasma contamination.

Bacterial growth

S. aureus strain LAC/USA300 was grown overnight in tryptic soy broth (TSB) with shaking at 37 °C and diluted 1/100 followed by an additional 3–4 h of growth until bacteria reached an optical density of 2. *S. pneumoniae* strain D39 was grown overnight in Luria–Bertani (LB) broth at 37 °C. The following day *S. pneumoniae* was diluted 1/50 followed by 4 h of growth until the optical density reached 0.5. *C. rodentium* and *S. Typhimurium* were grown overnight in LB broth at 37 °C. The following day, both were diluted 1/10 followed by 4 h of growth until the optical density reached 2.0. Bacterial density was confirmed by dilution plating. 1×10^9 CFU of each bacterial strain were boiled at 95 °C for 2 h and resuspended in PBS for experiments with heat-killed bacteria.

shRNA knockdown

Lentivirus-based knockdown of human *ATG16L1* (5′-CCGGACTG TAGCTTTGCCGTGAATGCTCGAGCATTCACGGCAAAGCTACAGTTTTT TTG-3′), *ULK1* (5′-CCGGGCCCTTTGCGTTATATTGTATCTCGAGATA CAATATAACGCAAAGGGCTTTTT-3′), *ATG5* (5′-CCGGGATTCATGGA ATTGAGCCAATCTCGAGATTGGCTCAATCCATGAATCTTTTTT-3′), *ATG7* (5′-CCGGGCTTTGGGATTGACACATTTCTCGAGAAATGTGTC AAATCCCAAAGCTTTTT-3′), *ADAM10* (5′-CCGGCCAGGTGGAATTACTTA ATTCTCGAAGAATTAAGTAATTCCTGGTTTTT-3′) and nontargeting control were performed using MISSION shRNA constructs (Sigma-Aldrich) as described³¹. Viruses expressing shRNAs were produced by DNA transfection via Lipofectamine 3000 (ThermoFisher). Successful knockdown was confirmed by western blot and/or reverse transcription (RT) with quantitative polymerase chain reaction (qPCR).

Flow cytometry

A549 cells were stained for surface markers ADAM10 (human SHM14) and EpCAM (human 9C4) using antibodies from BioLegend. A fixable live/dead stain from BioLegend was used to exclude dead cells. For profiling, exosome pellets were resuspended in 100 µl of PBS and were stained with a combination of CD9 (human HI9a, mouse MZ3), CD63 (human H5C6, mouse NVG-2), CD81 (human 5a6, mouse Eat-2), CCR5 (mouse HM-CCR5) and/or ADAM10 surface antibodies from BioLegend for 60 min at 4 °C. Exosome fractions were then stained with PKH67 (Sigma) as recommended by the manufacturer. Exosome fractions were washed in 40 ml PBS and ultracentrifuged again at 100,000g for 60 min. Washed exosome pellets were resuspended in PBS. For LysoSensor experiments, A549 cells were stimulated for 4 h (1 µM CpG-A, 10 nM bafilomycin, or medium alone). Cells were loaded with 1 µM LysoSensor Green DND-189 (Molecular Probes, Invitrogen) diluted in prewarmed medium and incubated for 15 min at 37 °C. Cells were collected with

1 ml 5 mM PBS-EDTA and washed once with FACS buffer (5% FCS, 1× PBS, 2 mM EDTA). Cells and exosomes were analysed using Beckman Coulter Cytoflex Cytometer. For FACS-assisted purification, exosome identification and isolation were performed as described³², and stained with a combination of CD81, CD63 and PKH67. Exosomes were sorted using the FACS ARIA IIu SORP cell sorter.

Western blotting

We collected 1×10^6 cells, washed them with PBS, and suspended them in RIPA buffer (Thermo-Scientific) containing 10× protease inhibitor (Santa Cruz Biotechnologies). Tissue homogenate was then pelleted twice at 10,000g for 10 min at 4 °C. Protein concentration in the supernatant was measured by Bradford assay and reduced using 4× Laemmli buffer containing β-mercaptoethanol at 95 °C for 5 min. For gel electrophoresis, 10–30 µg of protein was run at 120 V for 1 h using a 4–12% gradient protein gel (Thermo Fisher). Proteins were then transferred to an Immuno-Blot polyvinylidene fluoride (PVDF) membrane through Bio-Rad semi-dry transfer apparatus for 1 h, at 12 V constant. Membrane was incubated for 30 min with 5% non-fat dairy milk, and mouse anti-β-actin (Abcam) at 1/10,000, polyclonal rabbit anti-ADAM10 (Cell Signaling) at 1/2,500, mouse anti-ATG16L1 (MBL) at 1/1,000, monoclonal rabbit anti-CD9 (Cell Signaling) at 1/1,000, monoclonal mouse anti-CD81 (Cell Signaling) at 1/1,000, polyclonal rabbit anti-ARF6 (Cell Signaling) at 1/2,000, polyclonal rabbit anti-ASS1 (Abcam) at 1/1,000, polyclonal rabbit anti-HLA (Sigma) at 1/5,000, polyclonal rabbit anti-SQSTM1 (Cell Signaling) at 1/2,000, monoclonal rabbit anti-LC3 (Cell Signaling) at 1/2,000, or polyclonal rabbit anti-STX17 (Abcam) at 1/1,000 were probed overnight at 4 °C. Membranes were washed three times for 5 min and probed with secondary antibody rabbit-anti mouse LICOR IRDye 800CW and goat anti-rabbit LICOR IRDye 800CW 680 antibodies for 1 h at room temperature. After additional washing, protein was then detected with a LICOR Odyssey CLX imaging system.

Exosome isolation

Forty-eight hours before isolation, 1×10^7 A549 cells were plated in 150-mm tissue-culture dishes. At 24 h before isolation, approximately 35–40 ml of 10% Dulbecco’s modified Eagle medium (DMEM: 10% fetal bovine serum (FBS), 1% nonessential amino acids and 1% penicillin/streptomycin) was removed, and fresh 10% DMEM without supplements was added to each dish. On the day of exosome isolation, medium from each plate was removed and centrifuged once at 500g for 10 min, then centrifuged once at 10,000g for 10 min. Supernatants were passed through a 0.22-µm filter and finally ultracentrifuged at 100,000g for 90 min. Following ultracentrifugation, supernatant was discarded. The pellet remaining after ultracentrifugation was collected and used for downstream analysis. When assessing by western blot, normalization was performed by controlling for the number of cells seeded onto the plate, and input loading was confirmed by probing CD9 in the depleted cell fractions. For exosome isolation from bronchoalveolar lavage (BAL), mice were killed and the trachea exposed. Using a 0.5-inch blunt-nose needle, 1 ml of PBS was flushed into the lungs and removed three times. The exosome-isolation protocol was then performed on the remaining BAL fraction as described above.

α-Toxin and exosome treatment of cultured cells

To determine cell sensitivity to α-toxin, we seeded 3×10^4 A549 cells in 96-well plates and allowed them to attach overnight. Various concentrations of α-toxin were then added and incubated together for 3 h at 37 °C. We collected 50 ml of supernatant and measured cell death, either by LDH release (which indicates pore formation; Promega CytoTox-One Kit) or by metabolic activity via CellTiter (Promega catalogue number G3582). Total cytolysis was calculated according to the manufacturer’s instructions. For experiments in which the protective ability of exosomes was analysed, 3×10^4 A549 cells were seeded in 96-well plates and allowed to attach overnight. Exosome fractions

Article

were isolated from 35 ml of A549 culture supernatant as described above. The exosome fraction was mixed with $1 \mu\text{g ml}^{-1}$ α -toxin in PBS. The exosome/ α -toxin mixture was incubated at 37°C for 30 min and then added to plated A549 cells.

Heat-killed bacteria, bacterial components and inhibitors

We seeded 5×10^6 A549 cells and allowed them to attach overnight. The following day, cells were washed with PBS and new medium was added, including 5×10^6 CFU of heat-killed *S. aureus*, 2×10^5 CFU of heat-killed *S. pneumoniae*, 5×10^6 CFU of heat-killed *C. rodentium*, 5×10^6 CFU of heat-killed *S. Typhimurium*, $250 \mu\text{g ml}^{-1}$ lipoteichoic acid (LTA, Sigma, catalogue number L2515), $1 \mu\text{g ml}^{-1}$ lipopolysaccharide (LPS, InvivoGen, catalogue number tlr1-3pelps), $10 \mu\text{g ml}^{-1}$ peptidoglycan (PDG, Sigma, catalogue number 77140), $2 \mu\text{M}$ CpG DNA, $0.5 \mu\text{g ml}^{-1}$ Pam2CSK (InvivoGen, catalogue number tlr1-pm2s-1), $0.5 \mu\text{g ml}^{-1}$ Pam3CSK (InvivoGen, catalogue number tlr1-pms), $0.5 \mu\text{g ml}^{-1}$ *S. aureus* (SA) genomic DNA (gDNA, $0.5 \mu\text{g ml}^{-1}$), $0.5 \mu\text{g ml}^{-1}$ *S. aureus* RNA, $2 \mu\text{M}$ GW4869 (Sigma, catalogue number D1692) or 200 nM Torin-1 (Tocris, catalogue number 4247). After 4 h or 18 h, supernatants were removed from cultures and exosomes were collected as described above. Exosomes were identified and quantified using the aforementioned flow-cytometry protocol.

Infection and exosome treatment of mice

Donor mice received an intranasal treatment of heat-killed *S. aureus* to induce exosome production. After 4–6 h, mice were bled submandibularly and plasma was collected. The exosome fraction was collected as described above for A549 cells. Recipient wild-type or *Atg16L1*tm mice each received exosomes intraperitoneally isolated from 1 ml of plasma on day –1, day 0, and day +1 of infection in a final volume of 1 ml of PBS. Mice were intravenously infected with USA300 *S. aureus* on day 0, and were monitored daily for signs of morbidity.

α -Toxin purification from *S. aureus*

Primers VJT1391 (5'-GGGGG-AAGCTT-gtttgatatggaactcctgaatttttcg-3'; the underlined sequence is the HindIII site) and VJT1395 (5'-GATAA-GC TAGC-tta-GTGGTGGTGGTGGTGGT-attgtcatttcttc-3'; the underlined sequence is the NheI site) were used to amplify the promoter region of *hla* followed by the *hla* gene and polyhistidine tag ($6 \times \text{His}$ tag) from the genomic DNA of *S. aureus* strain Newman by PCR. The PCR product was then cloned into the pOS1 plasmid using the HindIII and NheI restriction sites to generate the pOS1-phla-hla-6his plasmid. The purified plasmid was transformed into *Escherichia coli* DH5 α competent cells, selected by ampicillin resistance ($100 \mu\text{g ml}^{-1}$) and confirmed by colony PCR and Sanger sequencing (Genewiz). The plasmid from a positive clone was purified and electroporated into *S. aureus* RN4220, selected by resistance to chloramphenicol ($10 \mu\text{g ml}^{-1}$); the plasmid purified from RN4220 was then electroporated into *S. aureus* Newman ΔlukED $\Delta\text{hlgACB}::\text{tet}$ $\Delta\text{lukAB}::\text{spec}$ $\Delta\text{hla}::\text{ermC}$ ($\Delta\Delta\Delta\Delta$) and selected for by resistance to chloramphenicol ($10 \mu\text{g ml}^{-1}$) resistance. For purification of His-tagged α -toxin, the *S. aureus* Newman $\Delta\Delta\Delta\Delta$ strain harbouring the pOS1-phla-hla-6his plasmid (strain VJT 45.56) were grown overnight in 5 ml TSB (Fisher) supplemented with chloramphenicol ($10 \mu\text{g ml}^{-1}$) at 37°C , shaking at 180 rpm, then subcultured the following day at a 1/100 dilution in TSB supplemented with chloramphenicol ($10 \mu\text{g ml}^{-1}$) and incubated for 5 h at 37°C , shaking at 180 rpm. The cultures were centrifuged for 15 min at 6,000 rpm and 4°C , and the supernatants were filter-sterilized through a $0.22\text{-}\mu\text{m}$ filter (Corning). The filtrates were incubated in the presence of a final concentration of 10 mM imidazole and nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) equilibrated with 10 mM imidazole (Fisher) in $1\times$ Tris-buffered saline (TBS; Cellgro) for 30 min at 4°C while nutating. The filtrates were passed through a glass column by gravity filtration, then Ni-NTA-bound toxins were washed with 25 mM imidazole, followed by a secondary wash with $1\times$ TBS. The Ni-NTA-bound toxins were eluted using 500 mM imidazole. The eluted toxins were dialysed into 10% glycerol in $1\times$ TBS

and filtered through a $0.22\text{-}\mu\text{m}$ filter before storage at -80°C . When required, the toxins were concentrated using concentrator columns (Ultra-15 Centrifugal Filter Units 10,000 NMWL, 15-ml volume capacity; EMD Millipore Amicon) before measuring protein concentration using absorbance at 280 nm with a Nanodrop (Thermo Scientific) and Beer-Lambert's equation. We separated $2 \mu\text{g}$ of the purified proteins by SDS-PAGE at 90 V for 120 min, followed by Coomassie blue staining to confirm protein purity by visualization.

Sample preparation for mass spectrometry

Exosomes were lysed in 8 M urea containing 10% SDS. Lysed exosomes were reduced using dithiothreitol ($5 \mu\text{l}$ of 0.2 M concentration) for 1 h at 55°C . The reduced cysteines were subsequently alkylated with iodoacetamide ($5 \mu\text{l}$ of 0.5 M) for 45 min in the dark at room temperature. Each sample was loaded onto S-trap microcolumns (Protifi) according to the manufacturer's instructions. In brief, $3 \mu\text{l}$ of 12% phosphoric acid and $165 \mu\text{l}$ of binding buffer (90% methanol, 100 mM triethylammonium bicarbonate (TEAB)) were added to each sample. Samples were loaded onto the S-trap columns and centrifuged at $4,000g$ for 30 s. After three washes, $20 \mu\text{l}$ of 50 mM TEAB and $1 \mu\text{g}$ of trypsin (1/50 ratio) were added to the trap and incubated at 47°C for 1 h. Peptides were then eluted using 40% acetonitrile (ACN) in 0.5% acetic acid followed by 80% ACN in 0.5% acetic acid. Eluted peptides were dried and concentrated in a SpeedVac.

Liquid chromatography–tandem mass spectrometry analysis

We loaded $1 \mu\text{g}$ of each sample onto a trap column (Acclaim PepMap 100 pre-column, $75 \mu\text{m} \times 2 \text{ cm}$, C18, $3 \mu\text{m}$, 100 \AA , Thermo Scientific) connected to an analytical column (EASY-Spray column, $50 \mu\text{m} \times 75 \mu\text{m}$ ID, PepMap RSLC C18, $2 \mu\text{m}$, 100 \AA , Thermo Scientific) using the autosampler of an Easy nLC1000 (Thermo Scientific) with solvent A consisting of 2% acetonitrile in 0.5% acetic acid and solvent B consisting of 80% acetonitrile in 0.5% acetic acid. The peptide mixture was gradient eluted into the Orbitrap QExactive HF-X Mass Spectrometer (Thermo Scientific) using the following gradient: 5–35% solvent B for 120 min, 35–45% solvent B for 10 min, and 45–100% solvent B for 20 min. The full scan was acquired with a resolution of 60,000 (at an m/z ratio of 200), a target value of 3×10^6 and a maximum ion time of 45 ms. Following each full mass-spectrometry (MS) scan, 20 data-dependent MS/MS spectra were acquired. These MS/MS spectra were collected with a resolution of 15,000, an AGC target of 1×10^5 , a maximum ion time of 120 ms, one microscan, a $2 m/z$ isolation window, a fixed first mass of $150 m/z$, a dynamic exclusion of 30 s, and a normalized collision energy of 27.

Analysis of mass-spectrometry data

All acquired MS/MS spectra were searched against the UniProt mouse reference database using Sequest HT within Proteome Discoverer 1.4 (Thermo Fisher Scientific). The parameters for searching MS/MS data were set as follows: precursor mass tolerance $\pm 10 \text{ ppm}$, fragment mass tolerance $\pm 0.02 \text{ Da}$, digestion enzyme trypsin allowing two missed cleavages, fixed modification of carbamidomethyl on cysteine, variable modification of oxidation on methionine, and variable modification of deamidation on glutamine and asparagine. The results were filtered using a 1% peptide and protein false discovery rate searched against a decoy database and requiring proteins to have at least two unique peptides.

α -Toxin oligomerization assay

Exosomes were collected from A549 culture supernatants as described above. Exosome fractions were resuspended in $30 \mu\text{l}$ PBS. α -Toxin was added to exosome suspension at a concentration of $1 \mu\text{g ml}^{-1}$. The exosome/ α -toxin combination was then shaken at room temperature for 1 h. Following incubation, the exosome/ α -toxin mixture was resuspended in 40 ml PBS and spun at $100,000g$ for 90 min to pellet exosomes with bound α -toxin and remove excess α -toxin. The exosome fraction was resuspended in RIPA buffer containing $4\times$ Laemmli buffer

without β -mercaptoethanol. For gel electrophoresis, each sample (including an α -toxin-only lane) was run at 120 V for 1 h using a 4–20% tris-glycine gradient protein gel (Thermo Fisher). Proteins were then transferred to an Immuno-Blot PVDF membrane through Bio-Rad semi-dry transfer apparatus for 1 h at a constant voltage of 12 V. Membrane was incubated for 30 min with 5% non-fat dairy milk, and mouse anti- α -toxin (Sigma) at 1/5,000 was probed overnight at 4 °C. Membranes were washed three times for 5 min and probed with secondary antibody goat anti-rabbit LICOR IRDye 800CW 680 antibodies for 1 h at room temperature. After additional washing, protein was then detected with a LICOR Odyssey CLX imaging system.

Transmission electron microscopy

For analysis of exosome morphology, we placed 5 μ l of isolated exosomes on glow-discharged carbon-coated 400-mesh copper/rhodium grids and stained the samples with 1% uranyl acetate aqueous solution. For whole-mount immune-electron microscopy, we deposited 5 μ l of 2%-paraformaldehyde-fixed exosomes on glow-discharged formvar-carbon-coated copper grids, and allowed the samples to adsorb for 20 min. After washing with PBS, the grids were incubated with 50 mM glycine/PBS for 5 min, blocked with 1% coldwater fish skin gelatin (Sigma) for 10 min, and incubated with primary antibodies (anti-TSG101, Abcam) in blocking solution for 2 h at room temperature. Following washing with PBS, gold-conjugated secondary antibodies (15 nm protein-A–gold, Cell Microscopy Center, University Medical Center Utrecht; 12-nm colloidal gold AffiniPure goat anti-rabbit IgG (H+L), Jackson ImmunoResearch Laboratories) were applied in the blocking buffer for 1 h. After washing with PBS, the grids were fixed in 1% glutaraldehyde in PBS for 5 min, washed with water, contrasted and embedded in a mixture of 3% uranyl acetate and 2% methylcellulose at a ratio of 1/9. All stained grids were examined under a Philips CM-12 electron microscope and photographed with a Gatan (4kx2.7k) digital camera (Gatan, Pleasanton, CA)³³.

Statistical analysis

All analyses were performed with Graphpad Prism v.7. The numbers of animals or biological replicates used herein were estimated on the basis of a power analysis with the following assumptions: the standard deviation will be roughly 20% of the mean; *P* values will be less than 0.05 when the null hypothesis is false; and the effect size (Cohen's *d*) is between 1.0 and 2.0. The minimal number of mice required under these conditions ranges from 6 to 28 for in vivo experiments. We have also carefully chosen the indicated sample size on the basis of empirical evidence of what is necessary to interpret the data and statistical significance. A unpaired two-tailed *t*-test with Welch's correction was used to evaluate differences between two groups. One-way ANOVA with Dunnett's post-test analysis was performed to evaluate differences between groups of three or more. The log-rank Mantel–Cox test was used for comparison of mortality curves. No randomization or blinding was used in this study.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The datasets generated herein are available from the corresponding authors upon request. Western blot gel source data can be found in Supplementary Fig. 1. All identified proteins from mass-spectrometry experiments and their accession identification codes are listed in Supplementary Tables 1–3. Source Data for Fig. 4 and Extended Data Fig. 6 are available with the paper.

29. Hubbard-Lucey, V. M. et al. Autophagy gene Atg16L1 prevents lethal T cell alloreactivity mediated by dendritic cells. *Immunity* **41**, 579–591 (2014).
30. Marchiando, A. M. et al. A deficiency in the autophagy gene Atg16L1 enhances resistance to enteric bacterial infection. *Cell Host Microbe* **14**, 216–224 (2013).
31. DuMont, A. L. et al. *Staphylococcus aureus* LukAB cytotoxin kills human neutrophils by targeting the CD11b subunit of the integrin Mac-1. *Proc. Natl Acad. Sci. USA* **110**, 10794–10799 (2013).
32. van der Vlist, E. J., Nolte-t Hoen, E. N., Stoorvogel, W., Arksteijn, G. J. & Wauben, M. H. Fluorescent labeling of nano-sized vesicles released by cells and subsequent quantitative and qualitative analysis by high-resolution flow cytometry. *Nat. Protocols* **7**, 1311–1326 (2012).
33. Théry, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* **30**, 3.22.21–3.22.29 (2006).

Acknowledgements We thank members of the Cadwell and Torres laboratory for constructive comments over the years, and particularly I. Iinov and E. Anderson for providing purified *S. aureus* DNA and assisting with the LukED protection experiments, respectively. This work was supported in part by US National Institute of Health (NIH) grants R01 AI121244 (to K.C. and V.J.T.); R01 HL123340, R01 DK093668, R01 DK103788, R01 AI130945 and R01 HL125816 (to K.C.); R01 AI099394 and R01 AI105129 (to V.J.T.); T32 AI007180 and F31 HL137304 (to M.D.K.); and pilot awards from the NYU CTSA grant UL1TR001445 from the National Center for Advancing Translational Sciences (NCATS) (to K.C.) and NYU Cancer Center grant P30CA016087 (to K.C.). Cell sorting/flow cytometry (NYU Langone Health's Cytometry and Cell Sorting Laboratory), mass spectrometry (The Proteomics Laboratory at NYU Langone Health) and electron-microscopy studies (The Microscopy Laboratory at NYU Langone Health) are supported in part by NYU Langone Health's Laura and Isaac Perlmutter Cancer Center Support (grant P30CA016087) from the National Cancer Institute Langone. This work was also supported by a Faculty Scholar grant from the Howard Hughes Medical Institute, the Merieux Institute, the Kenneth Rainin Foundation, the Crohn's & Colitis Foundation and the Stony Wold-Herbert Fund (all to K.C.). K.C. and V.J.T. are Burroughs Wellcome Fund Investigators in the Pathogenesis of Infectious Diseases.

Author contributions M.D.K., V.J.T. and K.C. formulated the original hypothesis and designed the study. M.D.K. performed the experiments and analysed the data. K.L.C. provided assistance with experiments that examined vesicle-trafficking events downstream of TLR9 activation. F.-X.L. provided assistance with TEM, and A.D. and B.M.U. performed and assisted in the analysis of mass-spectrometry experiments. K.T. generated the construct to express α -toxin, and purified and characterized the toxin. D.U. provided essential reagents and insight into the shRNA transduction experiments. M.D.K., V.J.T. and K.C. wrote the manuscript. All authors commented on the manuscript, data and conclusions.

Competing interests V.J.T. is an inventor on patents and patent applications filed by NYU, which are currently under commercial license to Janssen Biotech. K.C. has consulted for, or received an honorarium from, Puretech Health, Genentech and AbbVie; has received research support from Puretech Health, Pacific Biosciences and Pfizer; and has a provisional patent, US Patent application number 15/625,934.

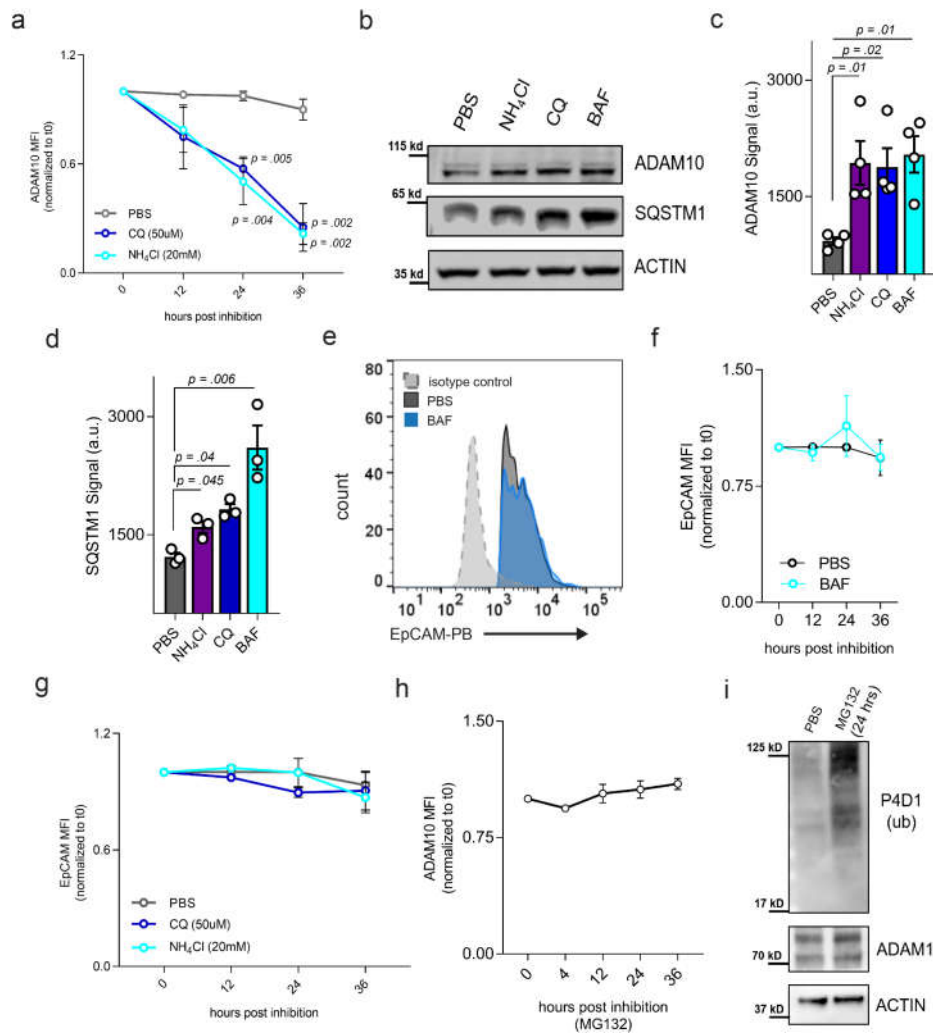
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2066-6>.

Correspondence and requests for materials should be addressed to V.J.T. or K.C.

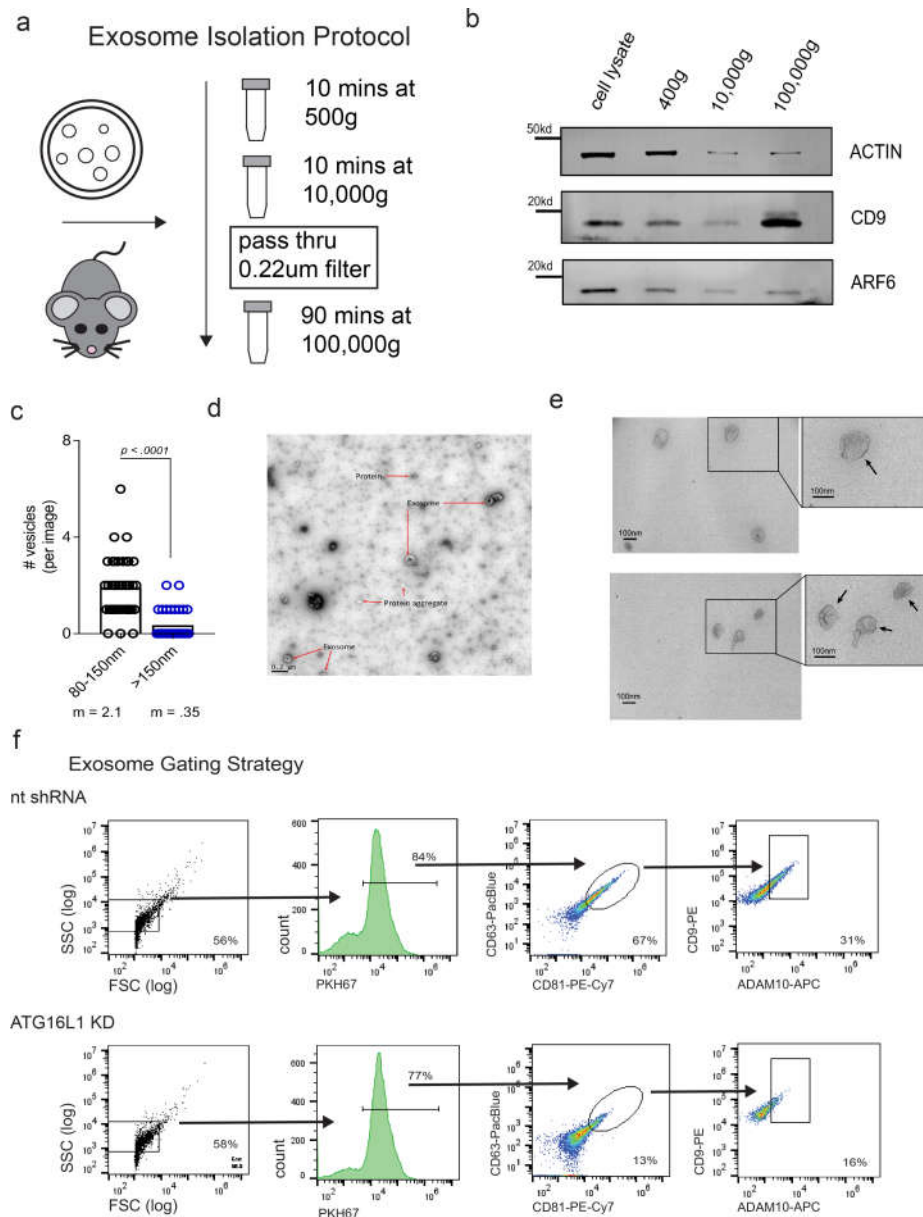
Peer review information Nature thanks Ivan Dikic, Lora Hooper, Andreas Peschel, Philip Stahl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | ADAM10 and EpCAM levels following lysosomal inhibition with ammonium chloride, chloroquine or bafilomycin or proteasomal inhibition with MG132. a, Time course of flow-cytometry analysis of ADAM10 following lysosomal inhibition with ammonium chloride (NH₄Cl, 20 mM), chloroquine (CQ, 50 μM) or PBS as a control; $n = 3$. **b–d**, Western blot analysis of ADAM10 and SQSTM1 following lysosomal inhibition with NH₄Cl, CQ or bafilomycin (BAF, 10 nM). Shown are a representative western blot from four independent experiments (**b**), quantification of ADAM10 levels ($n = 5$) (**c**) and quantification of SQSTM1 levels ($n = 3$) (**d**) at 24 h after inhibition. **e, f**, Representative histogram (**e**) and

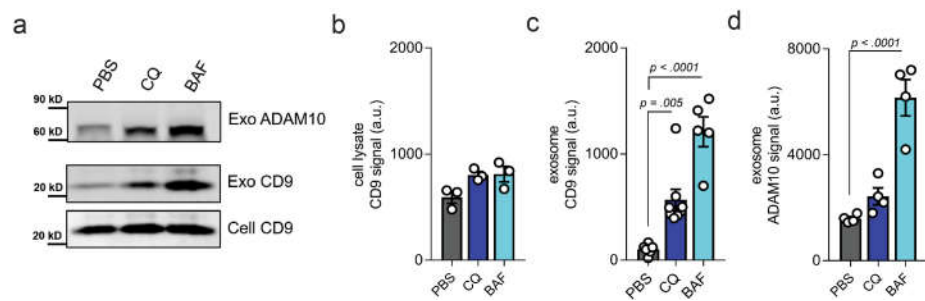
quantification (**f**) of cell-surface EpCAM in BAF-treated A549 cells; $n = 3$. **g**, Time course of flow-cytometry analysis of EpCAM following treatment with NH₄Cl or CQ; $n = 4$. **h, i**, ADAM10, P4D1 and actin levels following proteasomal inhibition with the chemical compound MG132. Shown are a flow-cytometry time course of cell-surface ADAM10 levels following MG132 treatment (**h**) and a representative western blot from three independent experiments (**i**); $n = 3$. Measurements were taken from distinct samples and graphs show mean \pm s.e.m. **a, c, d, f–h**, One-way ANOVA with Dunnett's post-test compared with PBS treatment or time 0.



Extended Data Fig. 2 | Exosome-isolation and quantification strategies.

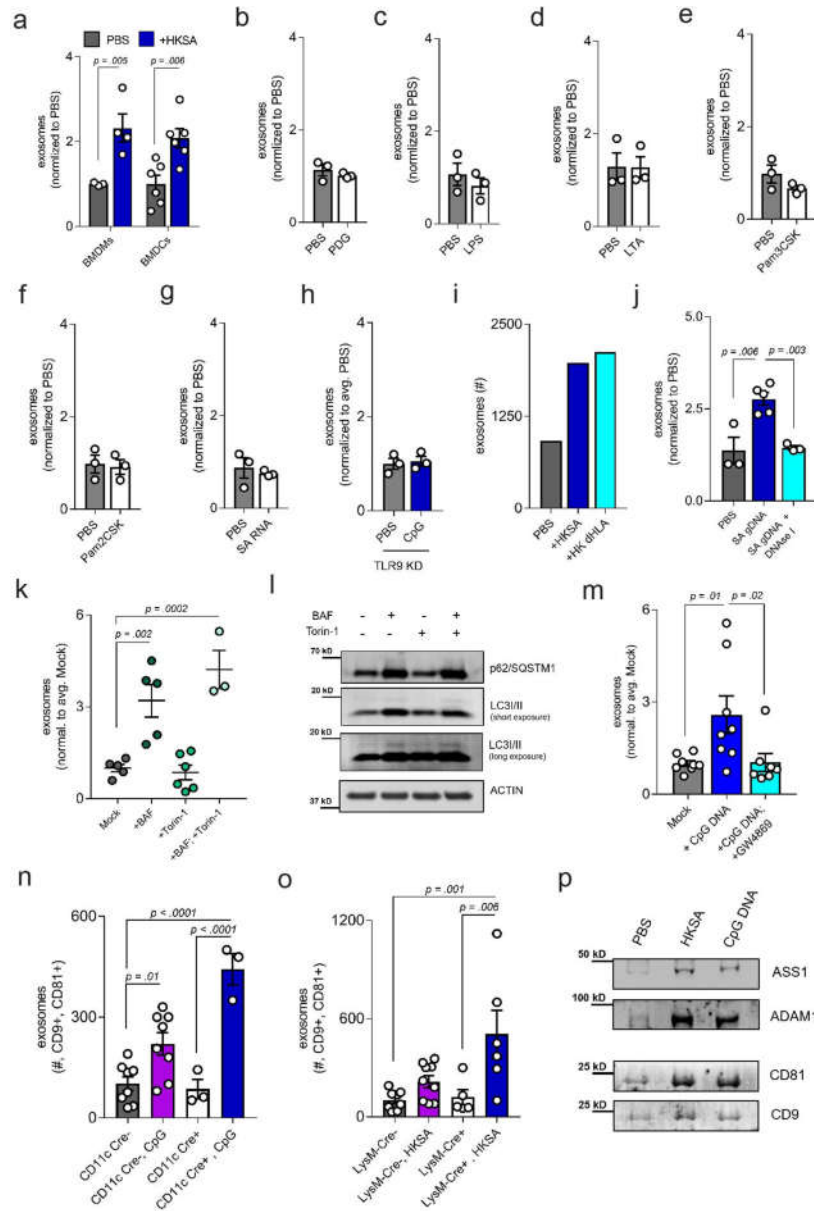
a, Exosome-isolation protocol from in vitro or in vivo sources. Exosomes are isolated using a multistep centrifugation procedure including a 0.22-µm filtration step. **b**, Western blot of actin, ARF6 and CD9 following each sequential centrifugation step during exosome isolation. **c**, Electron-microscopy (EM) quantification of vesicles 80–150 nm and greater than 150 nm in size; $n = 80$ images. **d**, EM negative staining of exosome fractions. Arrows indicate exosomes and protein aggregates. **e**, Representative EM images of the exosome fraction, and zoomed insets with arrows indicating the single

membranes of exosomes. **f**, Gating strategy and representative flow-cytometry plots from nt shRNA and ATG16L1KD samples of six independently repeated experiments. Exosomes were stained with antibodies against CD9, CD63, CD81 and ADAM10. Exosomes were concurrently labelled with PKH67, a lipid-membrane-incorporating dye. FSC, forward scatter; SSC, side scatter. Measurements were taken from distinct samples and graphs show means \pm s.e.m. **c**, Two-tailed, unpaired t -test with Welch's correction compared with PBS controls.



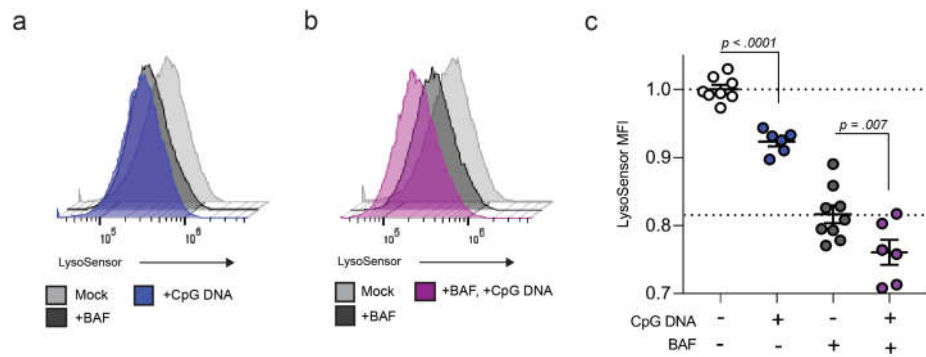
Extended Data Fig. 3 | CQ and BAF elicit production of ADAM10-positive exosomes. a–d, Western blot analysis of cell lysate CD9 (cell CD9), exosome CD9 (exo CD9) and exosome ADAM10 (exo ADAM10) following addition of CQ or BAF. Shown are a representative western blot from six independent

experiments (a) and quantification of cell CD9 (b), exosome CD9 (c), and exosome ADAM10 (d) after PBS, CQ or BAF treatment. Measurements were taken from distinct samples, and graphs show means ± s.e.m. b–d, One-way ANOVA with Dunnet’s post-test compared with PBS controls.



Extended Data Fig. 4 | Exosomes are produced in response to bacterial exposure. **a**, Flow-cytometry quantification of exosomes per 100,000 events in mouse BMDCs and BMDMs with or without exposure to HKSA (BMDCs plus PBS, $n = 6$; BMDCs plus HKSA, $n = 6$; BMDMs plus PBS, $n = 3$; BMDMs plus HKSA, $n = 4$). **b-h**, Quantification of total exosomes in A549 cell-culture supernatant by flow cytometry 18 h after treatment with peptidoglycan (PDG; **b**), lipopolysaccharide (LPS; **c**), lipoteichoic acid (LTA; **d**), Pam3CSK (**e**), Pam2CSK (**f**) or *S. aureus* RNA (SARNA; **g**); $n = 3$. **h**, Quantification of total exosomes in TLR9 KD A549 cell-culture supernatants following treatment with CpG DNA; $n = 3$. **i**, Flow-cytometry quantification of A549-produced exosomes following exposure to HKSA or to a strain of *S. aureus* deficient in the production of α -toxin (HK dHLA). **j**, Flow-cytometry quantification of exosomes isolated from A549 cells treated with PBS ($n = 3$), *S. aureus* genomic DNA (SAGDNA; $0.5 \mu\text{g ml}^{-1}$; $n = 5$), and/or DNaseI ($n = 2$). **k**, Flow-cytometry quantification of

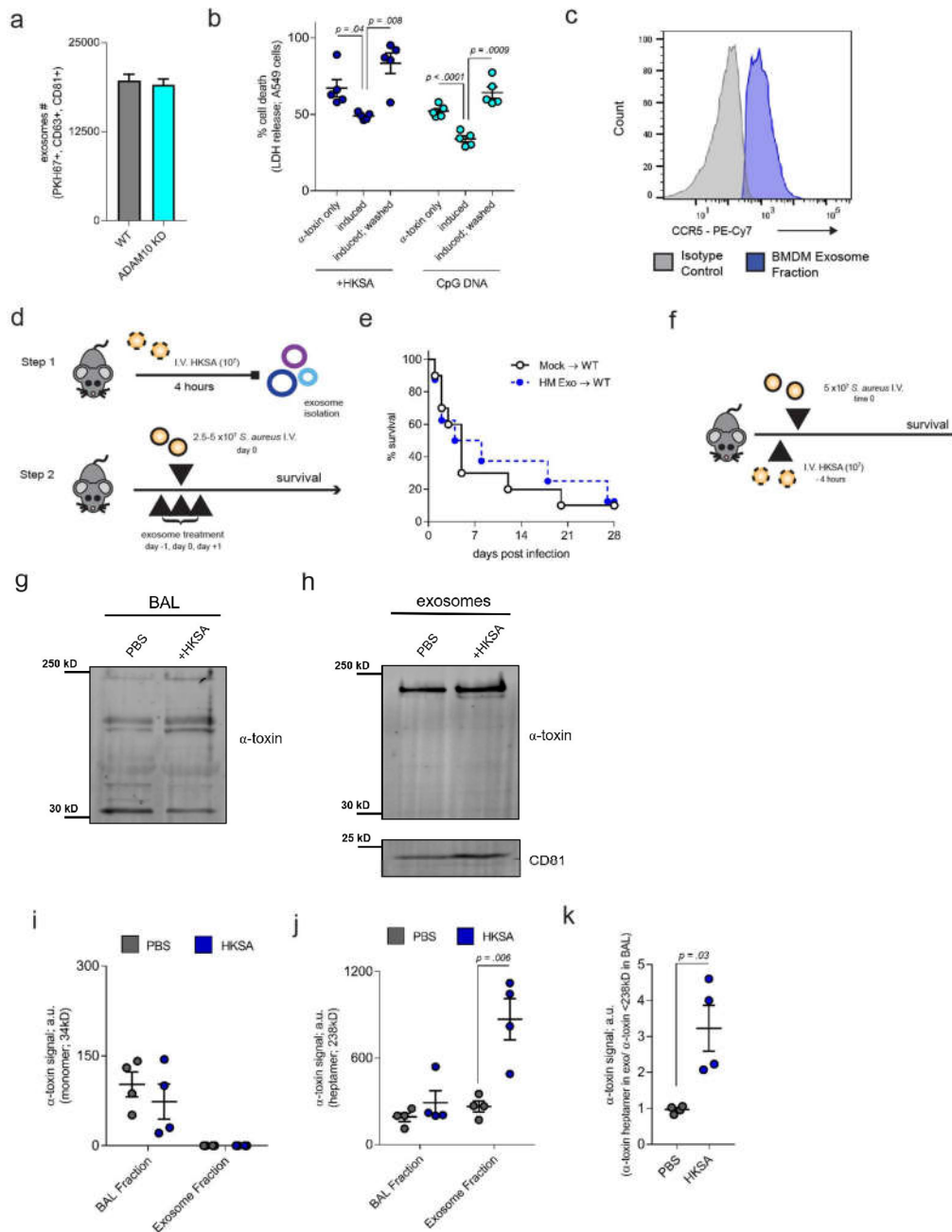
exosomes isolated from cells treated with BAF ($n = 5$), Torin-1 ($n = 6$), or both ($n = 3$). **l**, Representative western blot of SQSMT1, LC3/II and actin in cells treated with BAF, Torin-1 or both 4 h after treatment, from two independent experiments. **m**, Flow-cytometry quantification of exosomes from A549 cells treated with PBS (mock treatment; $n = 8$), CpG DNA ($n = 8$), or CpG DNA and GW4869 ($n = 7$). **n, o**, Plasma exosome quantification of ATG16L1 flow/flox; CD11c-Cre (**n**), and ATG16L1 flow/flox; LysM-Cre (**o**) following exposure to either CpG DNA or HKSA, respectively. **p**, Representative western blot of ADAM10, ASS1, CD9 and CD81 in exosome fractions submitted to mass spectrometry, from three independent experiments. Measurements were taken from distinct samples and graphs show means \pm s.e.m. **a, b-h**, Two-tailed, unpaired *t*-test with Welch's correction compared with PBS controls. **j, k, m-o**, One-way ANOVA with Dunnett's post-test compared with PBS, mock, CpG DNA or Cre-/- controls.



Extended Data Fig. 5 | BAF and CpG DNA decrease acidic organelles.

a–c, Representative flow-cytometry histograms from three independent experiments of Lysosensor signal following treatment with BAF or CpG DNA (**a**), or BAF with or without CpG DNA (**b**). **c**, Quantification of Lysosensor

MFI following treatment with PBS alone (no BAF or CpG DNA; $n = 8$), BAF ($n = 6$), CpG DNA ($n = 9$), or BAF plus CpG DNA ($n = 6$) (**c**). Measurements were taken from distinct samples and graphs show means \pm s.e.m. **c**, One-way ANOVA with Dunnet's post-test compared with PBS controls.



Extended Data Fig. 6 | Exosomes protect from *S. aureus* toxicity in vitro and in vivo. **a**, Flow-cytometry exosome quantification from nt shRNA control and ADAM10 KD A549 cells; $n = 3$. **b**, Cell death, measured by LDH release, of A549 cells treated with α -toxin only, pretreated with HKSA or CpG DNA and α -toxin ('induced'), or pre-exposed to HKSA or CpG DNA followed by PBS wash and then α -toxin treatment ('induced; washed'); $n = 5$. **c**, Representative flow-cytometry histograms of CCR5 on CD81-positive, CD63-positive and CD9-positive exosomes isolated from mouse BMDMs. **d**, Exogenous exosome-transfer protocol. In step 1, donor mice are pre-exposed to HKSA i.v. to induce exosome production. In step 2, exosomes from donor mice are injected intraperitoneally (i.p.) on day -1, day 0 and day +1 following lethal i.v. injection of *S. aureus*. **e**, Survival of wild-type mice infected i.v. with a lethal dose of 5×10^7 CFU of *S. aureus* (USA300) that were mock-treated ($n = 10$) or injected i.p. with

exosomes from *Atg16l1^{HM}* mice ($n = 8$). **f**, Endogenous exosome-protection protocol. Mice are i.v. injected with HKSA to induce exosome production. Four hours later, mice are infected with a lethal dose of *S. aureus* ($2.5-5 \times 10^7$). **g, h**, Western blot analysis of α -toxin oligomerization in total BAL or in exosome fraction in BAL of mice pre-exposed to HKSA or PBS intranasally (i.n.), representative of four independent experiments. **i, j**, Quantification of α -toxin monomer (i) and heptamer (j) in BAL and exosome fraction following pre-exposure; $n = 4$. **k**, Ratio of α -toxin heptamer in exosome fraction to total α -toxin signal in BAL; $n = 4$. Measurements were taken from distinct samples and graphs show means \pm s.e.m. **a, i-k**, Two-tailed, unpaired *t*-test with Welch's correction. **b**, One-way ANOVA with Dunnett's post-test compared with α -toxin only or 'induced' controls. **e**, log-rank Mantel-Cox test.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Flow cytometry data was collected and performed on a Beckman Coulter CytoFlex S.
Western blots were imaged and analyzed using the Li-COR Odyssey Clx.

Data analysis

All data was analyzed using Graphpad Prism v7. FlowJo v 10 was used to analyze all flow cytometry experiments. Western blots were analyzed using the Li-COR Odyssey Image Studio Ver 5.2. Genewiz was used to evaluate and confirm colony PCR and sequencing data. Mass spectrometry data was analyzed using Proteome Discoverer 1.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Authors can confirm that all relevant data are included in the paper and/or its supplementary information files

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The number of animals used in the experiments in this study is estimated based on a power analysis with the following assumptions: standard deviation will be ~20% of the mean, p-value will be under 0.05 when the null hypothesis is false, the effect size (Cohen's d) is between 1.0-2.0. The minimal number of mice required under these conditions ranges between 6-28 for in vivo experiments. Additionally, we have carefully chosen the sample size listed below based on empirical evidence of what is necessary for interpretation of the data and statistical significance.
Data exclusions	No data was excluded from the analysis.
Replication	In all experiments reported, each cell culture and animal experiment was attempted at least two times. In all experiments, no attempts at replication failed.
Randomization	Groups were established based off genotype and infection status. All other aspects were randomized.
Blinding	Investigator blinding was not used or necessary in this study as no quantification of subtle data or phenotypes were required.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials All unique materials used are readily available from the authors.

Antibodies

Antibodies used	Flow: CCR5 (mouse HM-CCR5, 1:200), EpCAM (human, 9C4, 1:500), CD81 (human, 5a6; mouse, Eat-2, 1:200), CD9 (human, HI9a; mouse, MZ3, 1:200), CD63 (human, H5C6; mouse, NVG-2, 1:200), ADAM10 (human, SHM14, 1:200) all purchased from BioLegend. Western Blot: CD9 (Cat# 13174S, Cell Signaling, 1:1000), CD81 (cat# ab109201, Abcam, 1:1000), ADAM10 (Cat#
-----------------	--

AB19026, Millipore, 1:2500), LC3 (Cat# PM026, MBL, 1:1000), SQSTM1 (Cat# P0067, Sigma, 1:4000), HLA/a-toxin (Cat# S75531, Sigma, 1:5000), ACTIN (Cat# A5441, Sigma, 1:10,000), STX17 (Cat# NBP1-93968, Novus Biologicals, 1:2500), ARF6 (Cat# 5740S, Cell Signaling, 1:1000), ASS1 (Cat# ab124465, Abcam, 1:1000).

Validation

All reported commercial antibody validation can be found on vendor websites.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Human lung epithelial purchased from ATCC (A549), and human embryonal kidney cells (293FT, ThermoFisher cat#R70007) were used for lentiviral packaging.
Authentication	ATCC Cell Line Authentication, Service Sanger Sequencing
Mycoplasma contamination	Confirmed negative for mycoplasma
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	This study uses mice from the wildtype (WT) and Atg16L1 hypomorph on the C57BL/6 background. Both male and female mice approximately 8-10 weeks of age are used.
Wild animals	Study does not involve wild animals.
Field-collected samples	Study does not involve field-collected samples.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Exosomes were collected from culture supernatants or plasma via high speed centrifugation prior to staining with flow cytometry antibodies (as described in manuscript). Cells isolated from in vitro assays were removed using 5mM EDTA and washed in PBS containing FBS prior to staining.
Instrument	Beckman Coulter CytoFlex S
Software	Flow data was collected with BD CytExpert Software and analyzed using FloJo v 10.
Cell population abundance	In sorting experiment, purity of samples was determined at the time of sorting by running a sorted sample on the FACSria. All samples tested were more than 95% pure.
Gating strategy	Cells were gated on FCS-A and SSC-A. Doublets were removed by gating FSC-A v FSC-H. Live cells were gated as Zombie-Dye negative. Cells were then gated on for ADAM10 or EpCAM. Exosomes were gated as PKH67+CD81+CD63+CD9+. Exosomes were also stained with ADAM10 unless otherwise mentioned. All gates were set using an antibody isotype control samples.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

A pneumonia outbreak associated with a new coronavirus of probable bat origin

<https://doi.org/10.1038/s41586-020-2012-7>

Received: 20 January 2020

Accepted: 29 January 2020

Published online: 3 February 2020

Open access

 Check for updates

Peng Zhou^{1,5}, Xing-Lou Yang^{1,5}, Xian-Guang Wang^{2,5}, Ben Hu¹, Lei Zhang¹, Wei Zhang¹, Hao-Rui Si^{1,3}, Yan Zhu¹, Bei Li¹, Chao-Lin Huang², Hui-Dong Chen², Jing Chen^{1,3}, Yun Luo^{1,3}, Hua Guo^{1,3}, Ren-Di Jiang^{1,3}, Mei-Qin Liu^{1,3}, Ying Chen^{1,3}, Xu-Rui Shen^{1,3}, Xi Wang^{1,3}, Xiao-Shuang Zheng^{1,3}, Kai Zhao^{1,3}, Quan-Jiao Chen¹, Fei Deng¹, Lin-Lin Liu⁴, Bing Yan¹, Fa-Xian Zhan⁴, Yan-Yi Wang¹, Geng-Fu Xiao¹ & Zheng-Li Shi^{1,5}✉

Since the outbreak of severe acute respiratory syndrome (SARS) 18 years ago, a large number of SARS-related coronaviruses (SARSr-CoVs) have been discovered in their natural reservoir host, bats^{1–4}. Previous studies have shown that some bat SARSr-CoVs have the potential to infect humans^{5–7}. Here we report the identification and characterization of a new coronavirus (2019-nCoV), which caused an epidemic of acute respiratory syndrome in humans in Wuhan, China. The epidemic, which started on 12 December 2019, had caused 2,794 laboratory-confirmed infections including 80 deaths by 26 January 2020. Full-length genome sequences were obtained from five patients at an early stage of the outbreak. The sequences are almost identical and share 79.6% sequence identity to SARS-CoV. Furthermore, we show that 2019-nCoV is 96% identical at the whole-genome level to a bat coronavirus. Pairwise protein sequence analysis of seven conserved non-structural proteins domains show that this virus belongs to the species of *SARSr-CoV*. In addition, 2019-nCoV virus isolated from the bronchoalveolar lavage fluid of a critically ill patient could be neutralized by sera from several patients. Notably, we confirmed that 2019-nCoV uses the same cell entry receptor—angiotensin converting enzyme II (ACE2)—as SARS-CoV.

Coronaviruses have caused two large-scale pandemics in the past two decades, SARS and Middle East respiratory syndrome (MERS)^{8,9}. It has generally been thought that SARSr-CoV—which is mainly found in bats—could cause a future disease outbreak^{10,11}. Here we report on a series of cases caused by an unidentified pneumonia disease outbreak in Wuhan, Hubei province, central China. This disease outbreak—which started from a local seafood market—has grown substantially to infect 2,761 people in China, is associated with 80 deaths and has led to the infection of 33 people in 10 additional countries as of 26 January 2020¹². Typical clinical symptoms of these patients are fever, dry cough, breathing difficulties (dyspnoea), headache and pneumonia. Disease onset may result in progressive respiratory failure owing to alveolar damage (as observed by transverse chest computerized-tomography images) and even death. The disease was determined to be caused by virus-induced pneumonia by clinicians according to clinical symptoms and other criteria, including a rise in body temperature, decreases in the number of lymphocytes and white blood cells (although levels of the latter were sometimes normal), new pulmonary infiltrates on chest radiography and no obvious improvement after treatment with antibiotics for three days. It appears that most of the early cases had contact history with the original seafood market; however, the disease has now progressed to be transmitted by human-to-human contact.

Samples from seven patients with severe pneumonia (six of whom are sellers or deliverymen from the seafood market), who were admitted to the intensive care unit of Wuhan Jin Yin-Tan Hospital at the beginning of the outbreak, were sent to the laboratory at the Wuhan Institute of Virology (WIV) for the diagnosis of the causative pathogen (Extended Data Table 1). As a laboratory investigating CoV, we first used pan-CoV PCR primers to test these samples¹³, given that the outbreak occurred in winter and in a market—the same environment as SARS infections. We found five samples to be PCR-positive for CoVs. One sample (WIV04), collected from the bronchoalveolar lavage fluid (BALF), was analysed by metagenomics analysis using next-generation sequencing to identify potential aetiological agents. Of the 10,038,758 total reads—of which 1,582 total reads were retained after filtering of reads from the human genome—1,378 (87.1%) sequences matched the sequence of SARSr-CoV (Fig. 1a). By de novo assembly and targeted PCR, we obtained a 29,891-base-pair CoV genome that shared 79.6% sequence identity to SARS-CoV BJ01 (GenBank accession number AY278488.2). High genome coverage was obtained by remapping the total reads to this genome (Extended Data Fig. 1). This sequence has been submitted to GISAID (<https://www.gisaid.org/>) (accession number EPI_ISL_402124). Following the name given by the World Health Organization (WHO), we tentatively call it novel coronavirus 2019 (2019-nCoV). Four more full-length genome sequences of 2019-nCoV (WIV02, WIV05, WIV06 and

¹CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, China. ²Wuhan Jin Yin-Tan Hospital, Wuhan, China. ³University of Chinese Academy of Sciences, Beijing, China. ⁴Hubei Provincial Center for Disease Control and Prevention, Wuhan, China. ⁵These authors contributed equally: Peng Zhou, Xing-Lou Yang, Xian-Guang Wang. ✉e-mail: zllshi@wh.iov.cn

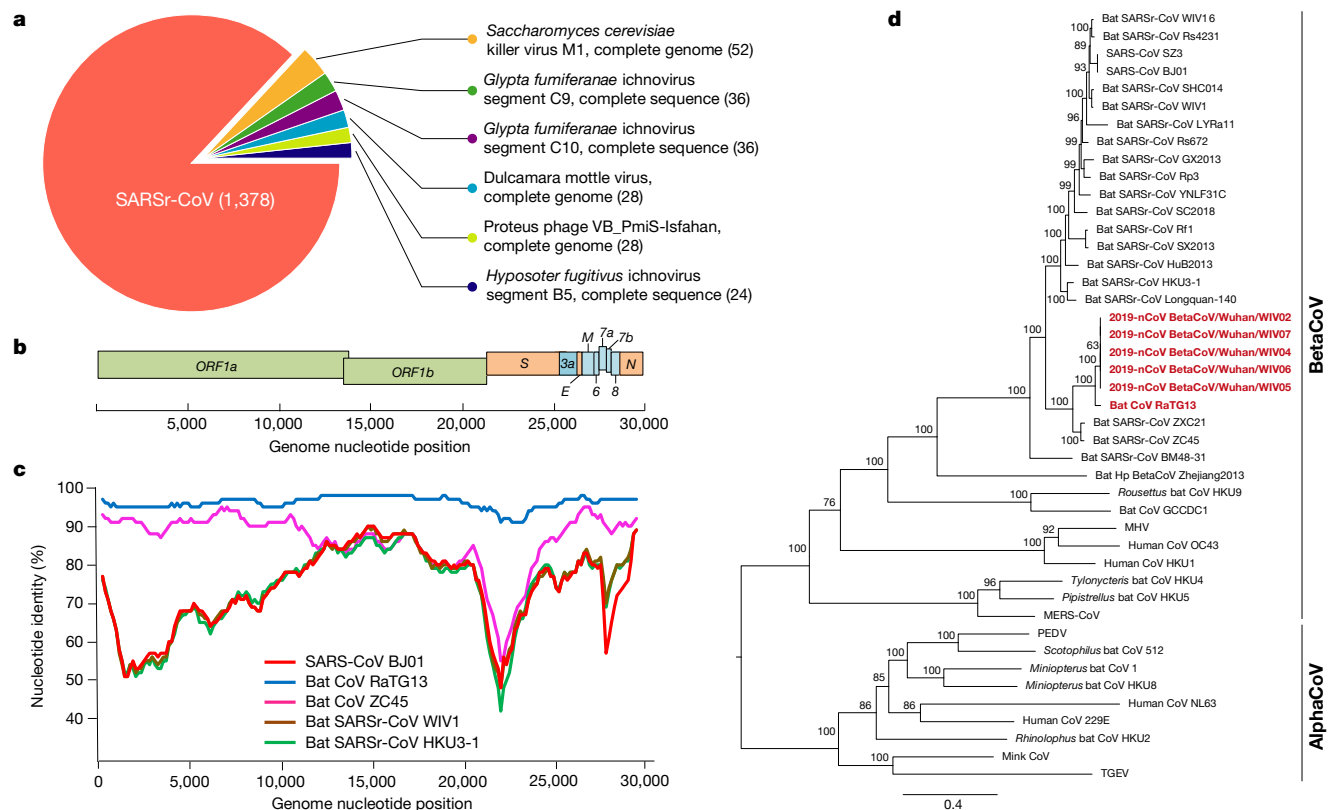


Fig. 1 | Genome characterization of 2019-nCoV. a, Metagenomics analysis of next-generation sequencing of BALF from patient ICU06. **b**, Genomic organization of 2019-nCoV WIV04. M, membrane. **c**, Similarity plot based on the full-length genome sequence of 2019-nCoV WIV04. Full-length genome sequences of SARS-CoV BJ01, bat SARSr-CoV WIV1, bat coronavirus RaTG13 and ZC45 were used as reference sequences. **d**, Phylogenetic tree based on

nucleotide sequences of complete genomes of coronaviruses. MHV, murine hepatitis virus; PEDV, porcine epidemic diarrhoea virus; TGEV, porcine transmissible gastroenteritis virus. The scale bars represent 0.1 substitutions per nucleotide position. Descriptions of the settings and software that was used are included in the Methods.

WIV07) (GISAID accession numbers EPI_ISL_402127–402130) that were more than 99.9% identical to each other were subsequently obtained from four additional patients using next-generation sequencing and PCR (Extended Data Table 2).

The virus genome consists of six major open-reading frames (ORFs) that are common to coronaviruses and a number of other accessory genes (Fig. 1b). Further analysis indicates that some of the 2019-nCoV genes shared less than 80% nucleotide sequence identity to SARS-CoV. However, the amino acid sequences of the seven conserved replicase domains in ORF1ab that were used for CoV species classification were 94.4% identical between 2019-nCoV and SARS-CoV, suggesting that the two viruses belong to the same species, *SARSr-CoV*.

We then found that a short region of RNA-dependent RNA polymerase (RdRp) from a bat coronavirus (BatCoV RaTG13)—which was previously detected in *Rhinolophus affinis* from Yunnan province—showed high sequence identity to 2019-nCoV. We carried out full-length sequencing on this RNA sample (GISAID accession number EPI_ISL_402131). Simplot analysis showed that 2019-nCoV was highly similar throughout the genome to RaTG13 (Fig. 1c), with an overall genome sequence identity of 96.2%. Using the aligned genome sequences of 2019-nCoV, RaTG13, SARS-CoV and previously reported bat SARSr-CoVs, no evidence for recombination events was detected in the genome of 2019-nCoV. Phylogenetic analysis of the full-length genome and the gene sequences of *RdRp* and *spike* (*S*) showed that—for all sequences—RaTG13 is the closest relative of 2019-nCoV and they form a distinct lineage from other SARSr-CoVs (Fig. 1d and Extended Data Fig. 2). The receptor-binding spike protein encoded by the *S* gene was highly divergent from other CoVs (Extended Data Fig. 2), with less than 75% nucleotide sequence

identity to all previously described SARSr-CoVs, except for a 93.1% nucleotide identity to RaTG13 (Extended Data Table 3). The *S* genes of 2019-nCoV and RaTG13 are longer than other SARSr-CoVs. The major differences in the sequence of the *S* gene of 2019-nCoV are the three short insertions in the N-terminal domain as well as changes in four out of five of the key residues in the receptor-binding motif compared with the sequence of SARS-CoV (Extended Data Fig. 3). Whether the insertions in the N-terminal domain of the *S* protein of 2019-nCoV confer sialic-acid-binding activity as it does in MERS-CoV needs to be further studied. The close phylogenetic relationship to RaTG13 provides evidence that 2019-nCoV may have originated in bats.

We rapidly developed a qPCR-based detection method on the basis of the sequence of the receptor-binding domain of the *S* gene, which was the most variable region of the genome (Fig. 1c). Our data show that the primers could differentiate 2019-nCoV from all other human coronaviruses including bat SARSr-CoV WIV1, which shares 95% identity with SARS-CoV (Extended Data Fig. 4a, b). Of the samples obtained from the seven patients, we found that six BALF and five oral swab samples were positive for 2019-nCoV during the first sampling, as assessed by qPCR and conventional PCR. However, we could no longer detect virus-positive samples in oral swabs, anal swabs and blood samples taken from these patients during the second sampling (Fig. 2a). However, we recommend that other qPCR targets, including the *RdRp* or *envelope* (*E*) genes are used for the routine detection of 2019-nCoV. On the basis of these findings, we propose that the disease could be transmitted by airborne transmission, although we cannot rule out other possible routes of transmission, as further investigation, including more patients, is required.

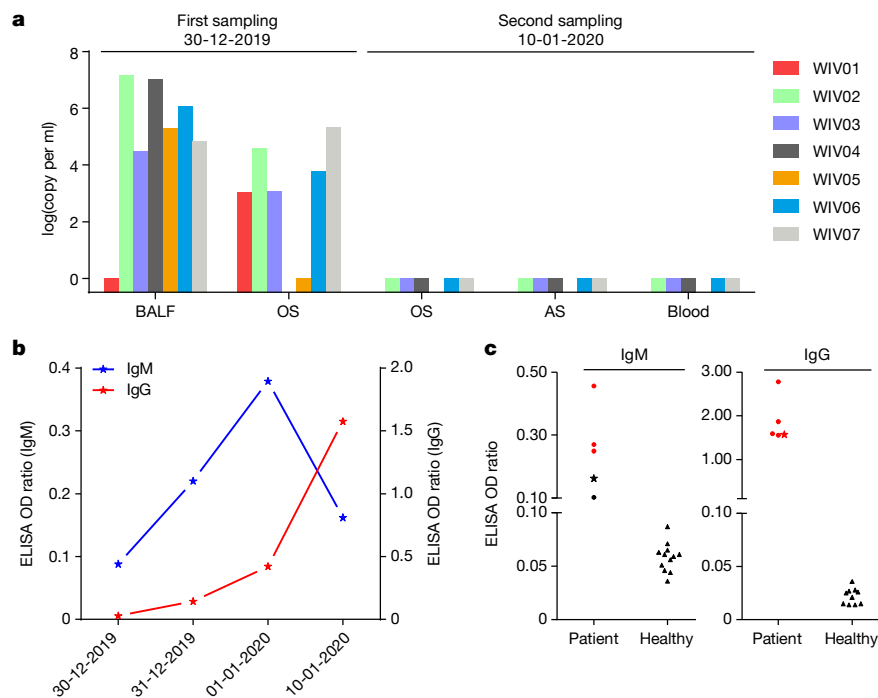


Fig. 2 | Molecular and serological investigation of patient samples.

a, Molecular detection of 2019-nCoV in seven patients. Patient information can be found in Extended Data Tables 1, 2. Detection methods are described in the Methods. AS, anal swab; OS, oral swab. **b**, Dynamics of 2019-nCoV antibody levels in one patient who showed signs of disease on 23 December 2019 (ICU-06). OD ratio, optical density at 450–630 nm. The right and left y axes indicate

ELISA OD ratios for IgM and IgG, respectively. **c**, Serological test of 2019-nCoV antibodies in five patients (Extended Data Table 2). The asterisk indicates data collected from patient ICU-06 on 10 January 2020. **b**, **c**, The cut-off was to 0.2 for the IgM analysis and to 0.3 for the IgG analysis, according to the levels of healthy controls.

For serological detection of 2019-nCoV, we used a previously developed nucleocapsid (N) protein from bat SARSr-CoV Rp3 as antigen for IgG and IgM enzyme-linked immunosorbent assays (ELISAs), as this protein shared 92% amino acid identity to N protein of 2019-nCoV (Extended Data Fig. 5) and showed no cross-reactivity against other human coronaviruses except SARSr-CoV⁷. We were only able to obtain five serum samples from the seven patients with viral infections. We monitored viral antibody levels in one patient (ICU-06) 7, 8, 9 and 18 days after the onset of disease (Extended Data Table 2). A clear trend was observed in the IgG and IgM titres, which increased over time, except that the IgM titre was decreased in the last sample (Fig. 2b). As a second analysis, we tested samples from 5 of the 7 virus-positive patients around 20 days after disease onset for the presence of viral antibodies (Extended Data Tables 1, 2). All patient samples—but not samples from healthy individuals—were strongly positive for viral IgG (Fig. 2b). There were also three IgM-positive samples, indicating an acute infection.

We next successfully isolated the virus (called 2019-nCoV BetaCoV/Wuhan/WIV04/2019) from both Vero E6 and Huh7 cells using the BALF sample of patient ICU-06. Clear cytopathogenic effects were observed in cells after incubation for three days (Extended Data Fig. 6a, b). The identity of the strain WIV04 was verified in Vero E6 cells by immunofluorescence microscopy using the cross-reactive viral N antibody (Extended Data Fig. 6c, d) and by metagenomics sequencing, most of the reads of which mapped to 2019-nCoV, and qPCR analysis showed that the viral load increased from day 1 to day 3 (Extended Data Fig. 6e, f). Viral particles in ultrathin sections of infected cells displayed a typical coronavirus morphology, as visualized by electron microscopy (Extended Data Fig. 6g). To further confirm the neutralization activity of the viral IgG-positive samples, we conducted serum-neutralization assays in Vero E6 cells using the five patient sera that were IgG-positive. We demonstrate that all samples were able to neutralize 100 TCID₅₀

(50% tissue-culture-infective dose) of 2019-nCoV at a dilution of 1:40–1:80. We also show that this virus could be cross-neutralized by horse anti-SARS-CoV serum (gift from L.-F. Wang) at dilutions of 1:40; however, the potential for cross-reactivity with SARS-CoV antibodies needs to be confirmed with anti-SARS-CoV serum from humans (Extended Data Table 4).

ACE2 is known to be a cell receptor for SARS-CoV¹⁴. To determine whether 2019-nCoV also uses ACE2 as a cellular entry receptor, we conducted virus infectivity studies using HeLa cells that expressed or did not express ACE2 proteins from humans, Chinese horseshoe bats, civets, pigs and mice. We show that 2019-nCoV is able to use all ACE2 proteins, except for mouse ACE2, as an entry receptor to enter ACE2-expressing cells, but not cells that did not express ACE2, indicating that ACE2 is probably the cell receptor through which 2019-nCoV enters cells (Fig. 3). We also show that 2019-nCoV does not use other coronavirus receptors, such as aminopeptidase N (APN) and dipeptidyl peptidase 4 (DPP4) (Extended Data Fig. 7).

The study provides a detailed report on 2019-nCoV, the likely aetiological agent responsible for the ongoing epidemic of acute respiratory syndrome in China and other countries. Virus-specific nucleotide-positive and viral-protein seroconversion was observed in all patients tested and provides evidence of an association between the disease and the presence of this virus. However, there are still many urgent questions that remain to be answered. The association between 2019-nCoV and the disease has not been verified by animal experiments to fulfil the Koch's postulates to establish a causative relationship between a microorganism and a disease. We do not yet know the transmission routine of this virus among hosts. It appears that the virus is becoming more transmissible between humans. We should closely monitor whether the virus continues to evolve to become more virulent. Owing to a shortage of specific treatments and considering the relatedness of 2019-nCoV to SARS-CoV, some drugs and pre-clinical vaccines against

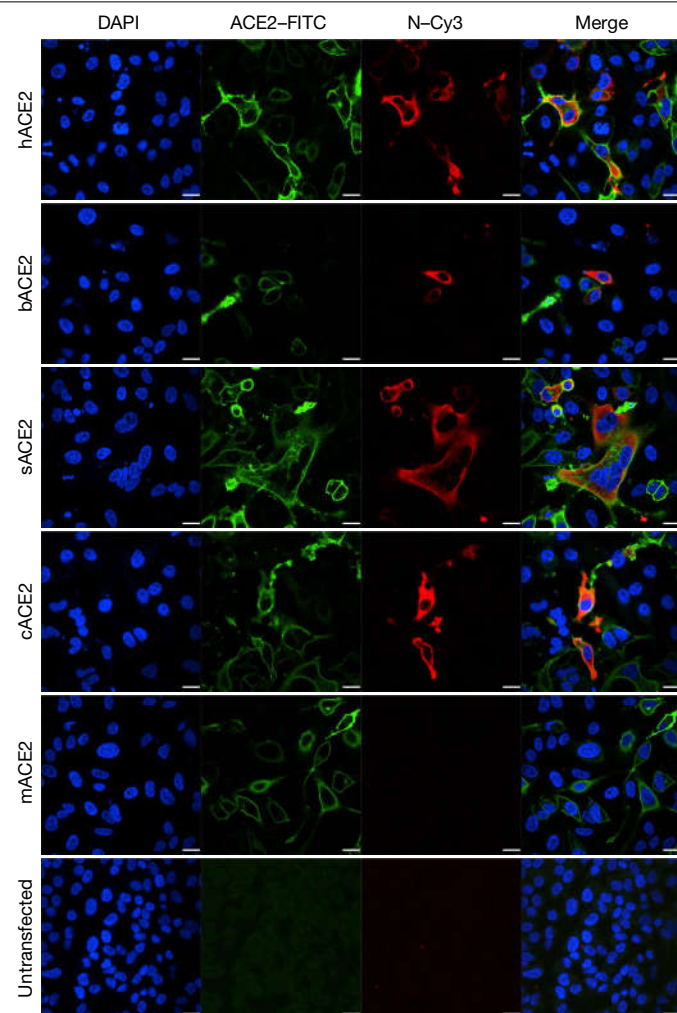


Fig. 3 | Analysis of the receptor use of 2019-nCoV. Determination of virus infectivity in HeLa cells that expressed or did not express (untransfected) ACE2. The expression of ACE2 plasmid with S tag was detected using mouse anti-S tag monoclonal antibody. hACE2, human ACE2; bACE2, ACE2 of *Rhinolophus sinicus* (bat); cACE2, civet ACE2; sACE2, swine ACE2 (pig); mACE2, mouse ACE2. Green, ACE2; red, viral protein (N); blue, DAPI (nuclei). Scale bars, 10 μ m.

SARS-CoV could probably be used to treat this virus. Finally, considering the wide spread of SARS-CoV in their natural reservoirs, future research should be focused on active surveillance of these viruses for broader geographical regions. In the long term, broad-spectrum antiviral drugs and vaccines should be prepared for emerging infectious diseases that are caused by this cluster of viruses in the future. Most importantly, strict regulations against the domestication and consumption of wildlife should be implemented.

Note added in proof: Since this paper was accepted, the ICTV has designated the virus as SARS-CoV-2¹⁵; in addition, the WHO has released the official name of the disease caused by this virus, which is COVID-19¹⁶.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1012-7>.

- Li, W. et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
- Ge, X.-Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
- Yang, L. et al. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerg. Infect. Dis.* **19**, 989–991 (2013).
- Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **13**, e1006698 (2017).
- Menachery, V. D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).
- Menachery, V. D. et al. SARS-like WIV1-CoV poised for human emergence. *Proc. Natl Acad. Sci. USA* **113**, 3048–3053 (2016).
- Wang, N. et al. Serological evidence of bat SARS-related coronavirus infection in humans, China. *Virology* **533**, 104–107 (2018).
- Drosten, C. et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
- Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
- Fan, Y., Zhao, K., Shi, Z.-L. & Zhou, P. Bat coronaviruses in China. *Viruses* **11**, 210 (2019).
- Wuhan Municipal Health Commission. Press statement related to novel coronavirus infection (in Chinese) <http://wjw.wuhan.gov.cn/front/web/showDetail/2020012709194> (2020).
- Poon, L. L. et al. Identification of a novel coronavirus in bats. *J. Virol.* **79**, 2001–2009 (2005).
- Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
- Gorbalenya, A. E. et al. Severe acute respiratory syndrome-related coronavirus — the species and its viruses, a statement of the Coronavirus Study Group. Preprint at <https://www.biorxiv.org/content/10.1101/2020.02.07.937862v1> (2020).
- WHO. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. <https://www.who.int/dg/speeches/detail/who-director-general-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020> (WHO, 11 February 2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sample collection

Human samples, including oral swabs, anal swabs, blood and BALF samples were collected by Jinyintan hospital (Wuhan, China) with the consent of all patients and approved by the ethics committee of the designated hospital for emerging infectious diseases. Patients were sampled without gender or age preference unless indicated. For swabs, 1.5 ml DMEM containing 2% FBS was added to each tube. The supernatant was collected after centrifugation at 2,500 rpm, vortexing for 60 s and a standing period of 15–30 min. The supernatant from swabs or BALF (no pre-treatment) was added to either lysis buffer for RNA extraction or to viral transport medium for isolation of the virus. The viral transport medium was composed of Hank's balanced salt solution (pH 7.4) containing BSA (1%), amphotericin (15 µg ml⁻¹), penicillin G (100 units ml⁻¹) and streptomycin (50 µg ml⁻¹). Serum was separated by centrifugation at 3,000g for 15 min within 24 h of collection, followed by inactivation at 56 °C for 1 h, and was then stored at 4 °C until use.

Virus isolation, cell infection, electron microscopy and neutralization assay

The following cell lines were used for virus isolation in this study: Vero E6 and Huh7 cells, which were cultured in DMEM containing 10% FBS. All cell lines were tested and free of mycoplasma contamination, submitted for species identification and authenticated by morphological evaluation by microscopy. None of the cell lines was on the list of commonly misidentified cell lines (by ICLAC).

Cultured cell monolayers were maintained in their respective medium. The PCR-positive BALF sample from ICU-06 patient was spun at 8,000g for 15 min, filtered and diluted 1:2 with DMEM supplemented with 16 µg ml⁻¹ trypsin before it was added to the cells. After incubation at 37 °C for 1 h, the inoculum was removed and replaced with fresh culture medium containing antibiotics (see below) and 16 µg ml⁻¹ trypsin. The cells were incubated at 37 °C and observed daily for cytopathogenic effects. The culture supernatant was examined for the presence of virus by qRT-PCR methods developed in this study, and cells were examined by immunofluorescence microscopy using the anti-SARSr-CoV Rp3 N antibody that was generated in-house (1:1,000). Penicillin (100 units ml⁻¹) and streptomycin (15 µg ml⁻¹) were included in all tissue culture media.

Vero E6 cells were infected with the new virus at a multiplicity of infection (MOI) of 0.5 and collected 48 h after infection. Cells were fixed with 2.5% (w/v) glutaraldehyde and 1% osmium tetroxide, dehydrated through a graded series of ethanol concentrations (from 30 to 100%) and embedded with epoxy resin. Ultrathin sections (80 nm) of embedded cells were prepared, deposited onto Formvar-coated copper grids (200 mesh), stained with uranyl acetate and lead citrate, and analysed using a 200-kV Tecnai G2 electron microscope.

The virus neutralization test was carried out in a 96-well plate. The patient serum samples were heat-inactivated by incubation at 56 °C for 1 h before use. The serum samples were diluted to 1:10, 1:20, 1:40 or 1:80, and then an equal volume of virus stock was added and incubated at 37 °C for 60 min in a 5% CO₂ incubator. Diluted horse anti-SARS-CoV serum or serum samples from healthy individuals were used as control. After incubation, 100 µl mixtures were inoculated onto a monolayer of Vero E6 cells in a 96-well plate for 1 h. Each serum was assessed in triplicate. After removing the supernatant, the plate was washed twice with DMEM medium. Cells were incubated with DMEM supplemented with 2% FBS for 3 days. Subsequently, the cells were checked for cytopathogenic effects.

RNA extraction and PCR

Whenever commercial kits were used, the manufacturer's instructions were followed without modification. RNA was extracted from 200 µl of samples with the High Pure Viral RNA kit (Roche). RNA was eluted in 50 µl of elution buffer and used as the template for RT-PCR.

For qPCR analysis, primers based on the S gene of 2019-nCoV were designed: RBD-qF1, 5'-CAATGGTTTAAACAGGCACAGG-3'; RBD-qR1, 5'-CTCAAGTGTCTGTGGATCAGC-3'. RNA extracted as described above was used for qPCR using the HiScript II One Step qRT-PCR SYBR Green Kit (Vazyme Biotech). Conventional PCRs were also performed using the following primer pairs: ND-CoVs-951F, 5'-TGT-KAGRTTYCCTAAYATTAC-3'; ND-CoVs-1805R, 5'-ACATCYTGATAN-ARAACAGC-3'. The 20-µl qPCR reaction mix contained 10 µl 2× One Step SYBR Green mix, 1 µl One Step SYBR Green Enzyme mix, 0.4 µl 50× ROX Reference Dye 1, 0.4 µl of each primer (10 µM) and 2 µl template RNA. Amplification was performed as follows: 50 °C for 3 min, 95 °C for 30 s followed by 40 cycles consisting of 95 °C for 10 s and 60 °C for 30 s, and a default melting curve step in an ABI 7500 Real-time PCR machine.

Serological test

In-house anti-SARSr-CoV IgG and IgM ELISA kits were developed using SARSr-CoV Rp3 N protein as antigen, which shared more than 90% amino acid identity to all SARSr-CoVs². For IgG analyses, MaxiSorp Nunc-immuno 96-well ELISA plates were coated (100 ng per well) overnight with recombinant N protein. Human sera were used at a dilution of 1:20 for 1 h at 37 °C. An anti-human IgG HRP-conjugated monoclonal antibody (Kyab Biotech) was used at a dilution of 1:40,000. The OD value (450–630 nm) was calculated. For IgM analyses, MaxiSorp Nunc-immuno 96-well ELISA plates were coated (500 ng per well) overnight with anti-human IgM (µ chain). Human sera were used at a 1:100 dilution for 40 min at 37 °C, followed by incubation with an anti-Rp3 N HRP-conjugated antibody (Kyab Biotech) at a dilution of 1:4,000. The OD value (450–630 nm) was calculated.

Examination of ACE2 receptor for 2019-nCoV infection

HeLa cells transiently expressing ACE2 were prepared using Lipofectamine 3000 (Thermo Fisher Scientific) in a 96-well plate; mock-transfected cells were used as controls. 2019-nCoV grown in Vero E6 cells was used for infection at a MOI of 0.5. APN and DPP4 were analysed in the same way. The inoculum was removed after absorption for 1 h and washed twice with PBS and supplemented with medium. At 24 h after infection, cells were washed with PBS and fixed with 4% formaldehyde in PBS (pH 7.4) for 20 min at room temperature. ACE2 expression was detected using a mouse anti-S tag monoclonal antibody and a FITC-labelled goat anti-mouse IgG H&L (Abcam, ab96879). Viral replication was detected using a rabbit antibody against the Rp3 N protein (generated in-house, 1:1,000) and a Cy3-conjugated goat anti-rabbit IgG (1:200, Abcam, ab6939). Nuclei were stained with DAPI (Beyotime). Staining patterns were examined using confocal microscopy on a FV1200 microscope (Olympus).

High-throughput sequencing, pathogen screening and genome assembly

Samples from patient BALF or from the supernatant of virus cultures were used for RNA extraction and next-generation sequencing (NGS) using BGI MGISEQ2000 and Illumina MiSeq 3000 sequencers. Metagenomic analysis was carried out mainly based on the bioinformatics platform MGmapper (PE_2.24 and SE_2.24). The raw NGS reads were first processed by Cutadapt (v.1.18) with minimum read length of 30 base pairs. BWA (v.0.7.12-r1039) was used to align reads to a local database with a filter hits parameter of 0.8 FMM ((match + mismatch)/read length ≥ fraction) value and minimum alignment score of 30. Parameters for post-processing of assigned reads were set to a minimum size normalized abundance of 0.01,

minimum read count of 20 and were otherwise set to default parameters. A local nucleic acid database for human and mammals was used to filter reads of host genomes before mapping reads to the virus database. The results of the metagenomic analysis were displayed as pie charts using Microsoft Office 2010. NGS reads were assembled into genomes using Geneious (v.11.0.3) and MEGAHIT (v.1.2.9). PCR and Sanger sequencing was performed to fill gaps in the genome. 5'-rapid amplification of cDNA ends (RACE) was performed to determine the 5'-end of the genomes using a SMARTer RACE 5'/3' kit (Takara). Genomes were annotated using the Clone Manager Professional Suite 8 (Sci-Ed Software).

Phylogenetic analysis

Routine sequence management and analysis was carried out using DNASTar. The sequence alignment of complete genome sequences was performed using MAFFT (v.7.307) with default parameters. The codon alignments of full-length S and RdRp gene sequences were converted from the corresponding protein alignments by PAL2NAL (v.14); the protein alignments were created by Clustal Omega (v.1.2.4) using default parameters. Maximum likelihood phylogenetic trees were generated using RAxML (v.0.9.0) with GTR+G substitution model and 1,000 bootstrap replicates.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence data that support the findings of this study have been deposited in GISAID (<https://www.gisaid.org/>) with accession numbers EPI_ISL_402124, EPI_ISL_402127–EPI_ISL_402130 and EPI_ISL_402131; GenBank with accession numbers MN996527–MN996532; National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences (<https://bigd.big.ac.cn/databases?lang=en>) with accession numbers SAMC133236–SAMC133240 and SAMC133252.

Acknowledgements We thank P. Zhang and A. Du from the WIV core facility centre for their help with producing transmission electron microscopy micrographs; H.-Z. Liu and P. Yu from WIV for bioinformatics analysis. This work was jointly supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB29010101 to Z.-L.S. and XDB29010104 to P.Z.), China Natural Science Foundation for excellent scholars (81822028 to P.Z., 31770175 to Z.-L.S. and 31800142 to B.H.), Mega-Project for Infectious Disease from Minister of Science and Technology of the People's Republic of China (2018ZX10305409-004-001 to P.Z.), Youth innovation promotion association of CAS (2019328 to X.-L.Y.).

Author contributions Z.-L.S., P.Z., Y.-Y.W. and G.-F.X. conceived the study. X.-G.W., C.-L.H., H.-D.C., F.D., Q.-J.C., F.-X.Z. and L.-L.L. collected patient samples. X.-L.Y., B.Y., W.Z., B.L., J.C., X.-S.Z., Y.L., H.G., R.-D.J., M.-Q.L., Y.C., X.W., X.-R.S. and K.Z. performed qPCR, serology and virus culturing experiments. L.Z., Y.Z., H.-R.S. and B.H. performed genome sequencing and annotations.

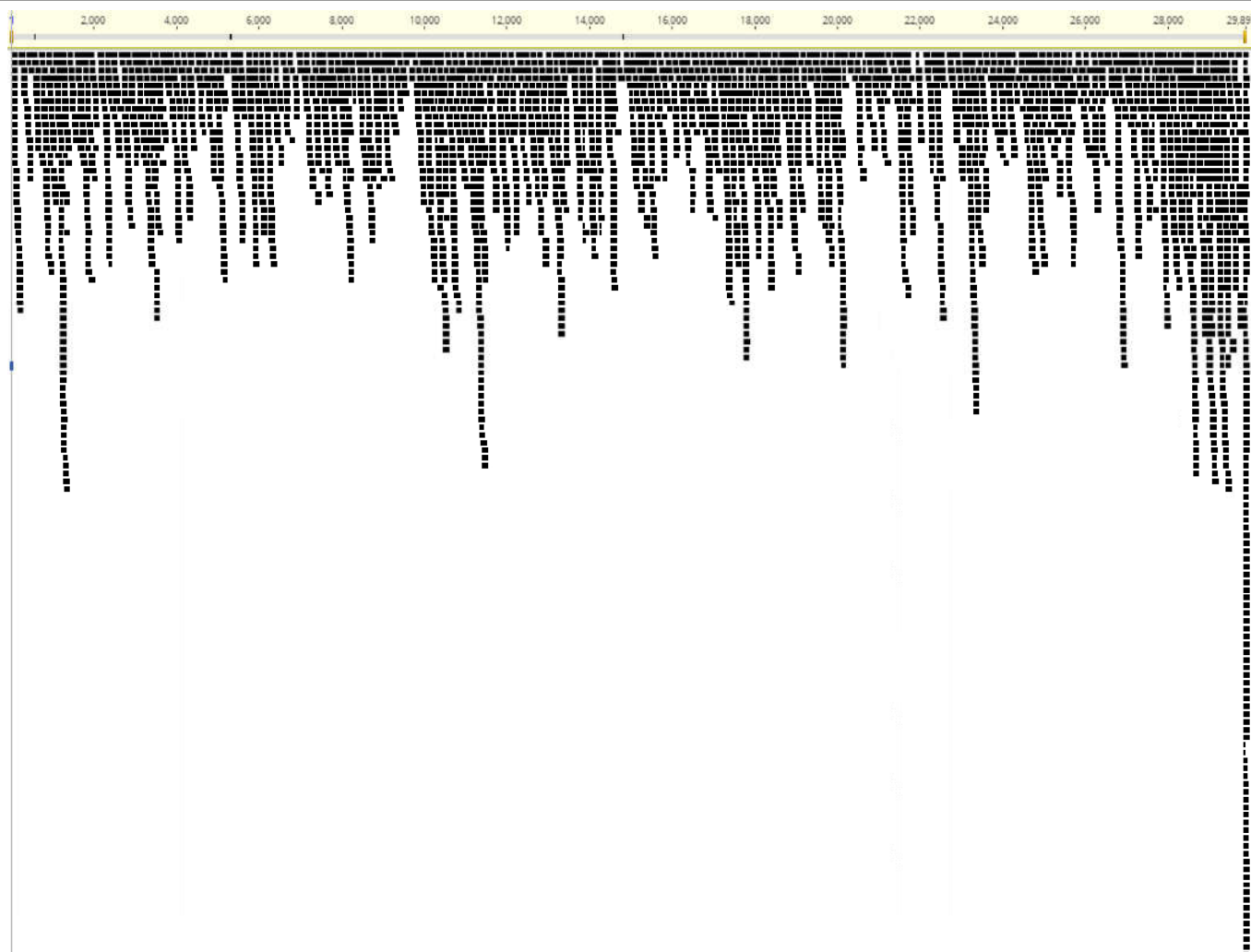
Competing interests The authors declare no competing interests.

Additional information

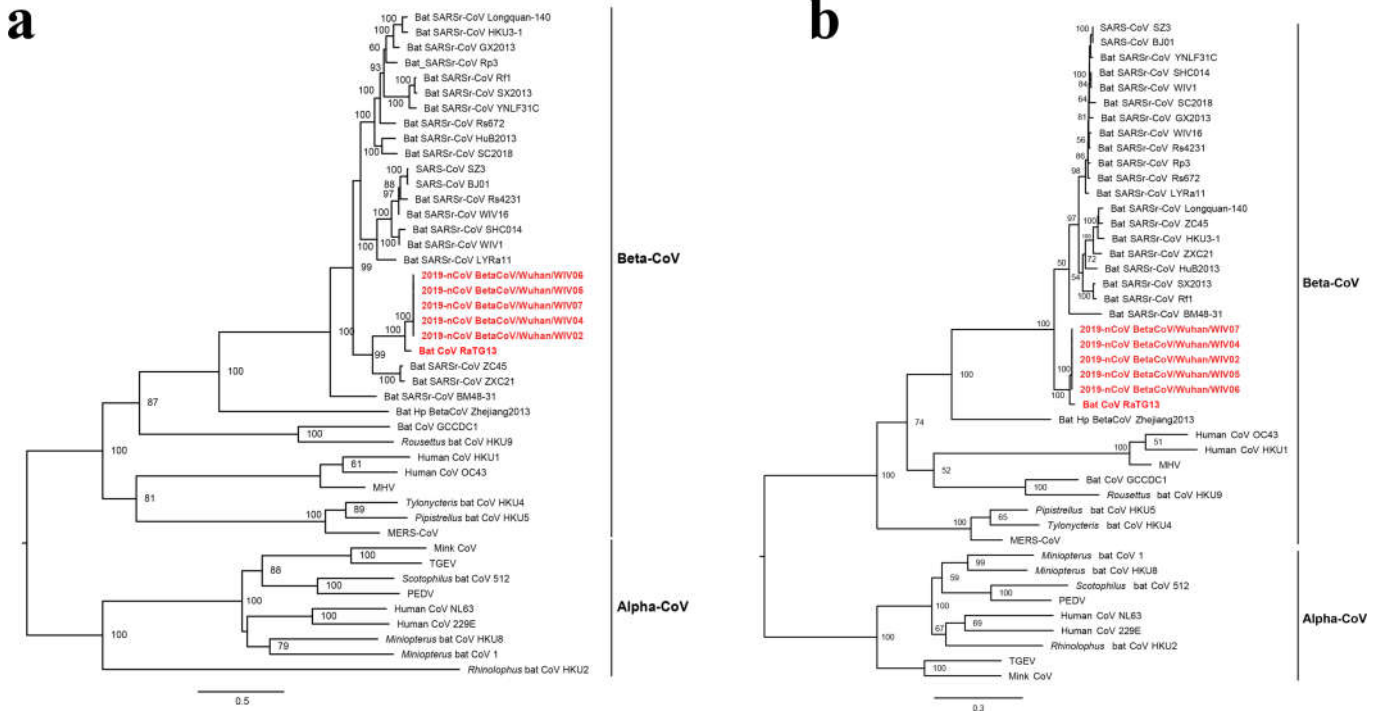
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2012-7>.

Correspondence and requests for materials should be addressed to Z.-L.S.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

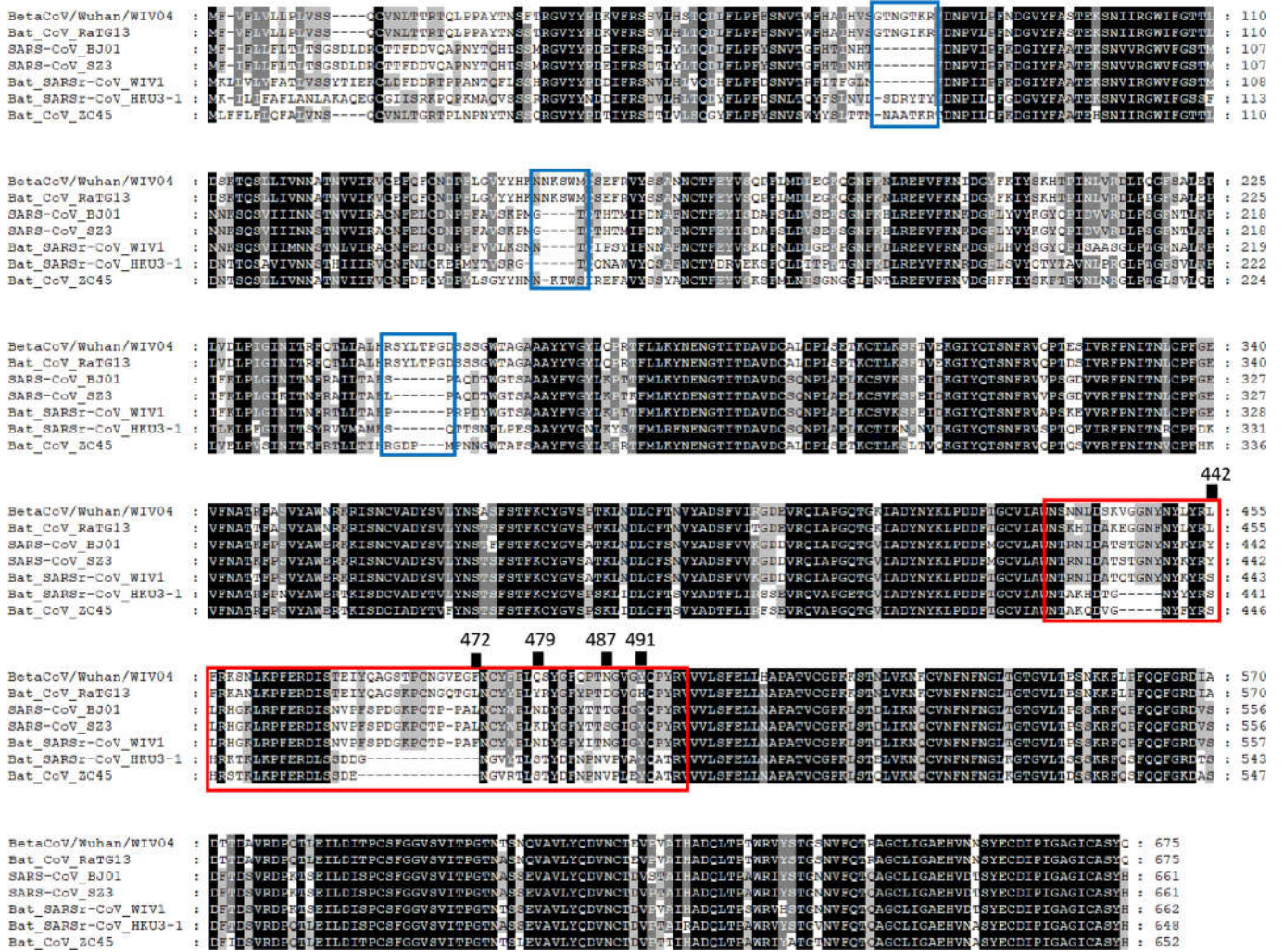


Extended Data Fig. 1 | NGS raw reads of sample WIV04 mapping to the 2019-nCoV sequence. The x-axis indicates the genome nucleotide position and the y-axis represents the read depth of the mapping.



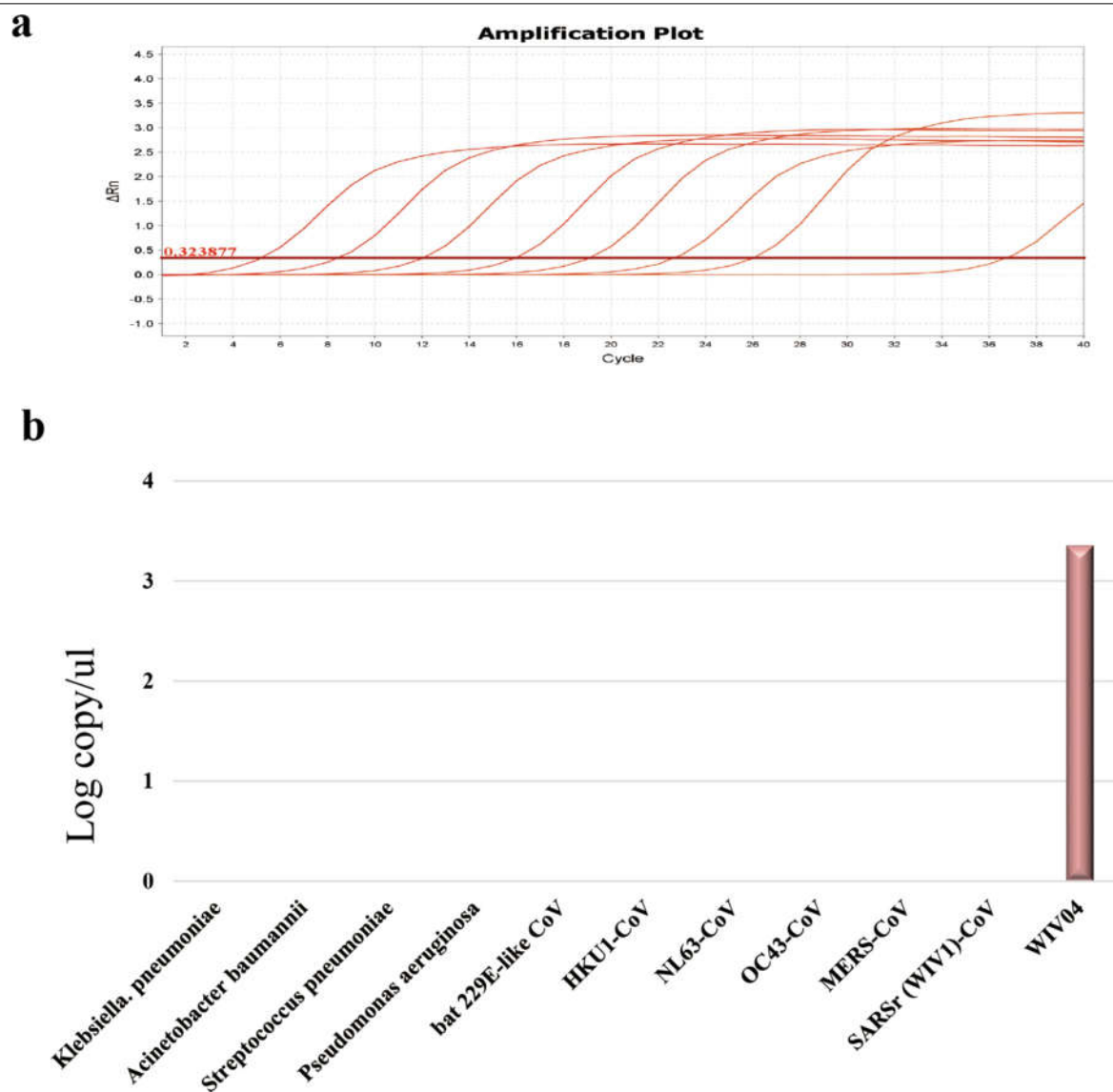
Extended Data Fig. 2 | Phylogenetic trees based on the complete S and RdRp gene sequences of coronaviruses. a, b. Phylogenetic trees on the basis of the gene sequences of S (a) and RdRp (b) are shown. 2019-nCoV and bat CoV RaTG13 are shown in bold and in red. The trees were constructed using the maximum

likelihood method using the GTR + G substitution model with bootstrap values determined by 1,000 replicates. Bootstrap values of more than 50% are shown.



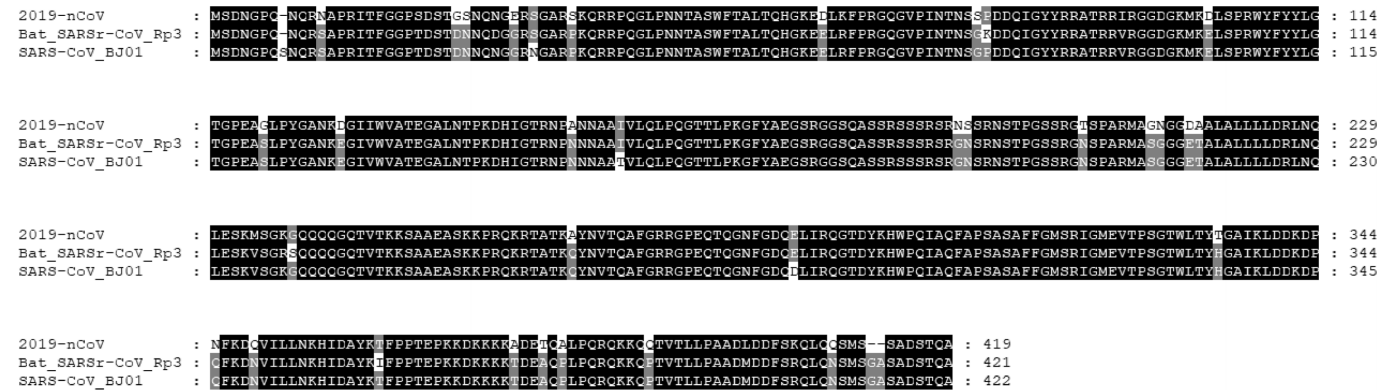
Extended Data Fig. 3 | Amino acid sequence alignment of the S1 protein of the 2019-nCoV to SARS-CoV and selected bat SARSr-CoVs. The receptor-binding motif of SARS-CoV and the homologous region of other coronaviruses are indicated by the red box. The key amino acid residues involved in the interaction with human ACE2 are numbered at the top of the aligned

sequences. The short insertions in the N-terminal domain of the 2019-nCoV are indicated by the blue boxes. Bat CoV RaTG13 was obtained from *R. affinis*, found in Yunnan province. Bat CoV ZC45 was obtained from *R. sinicus*, found in Zhejiang province.

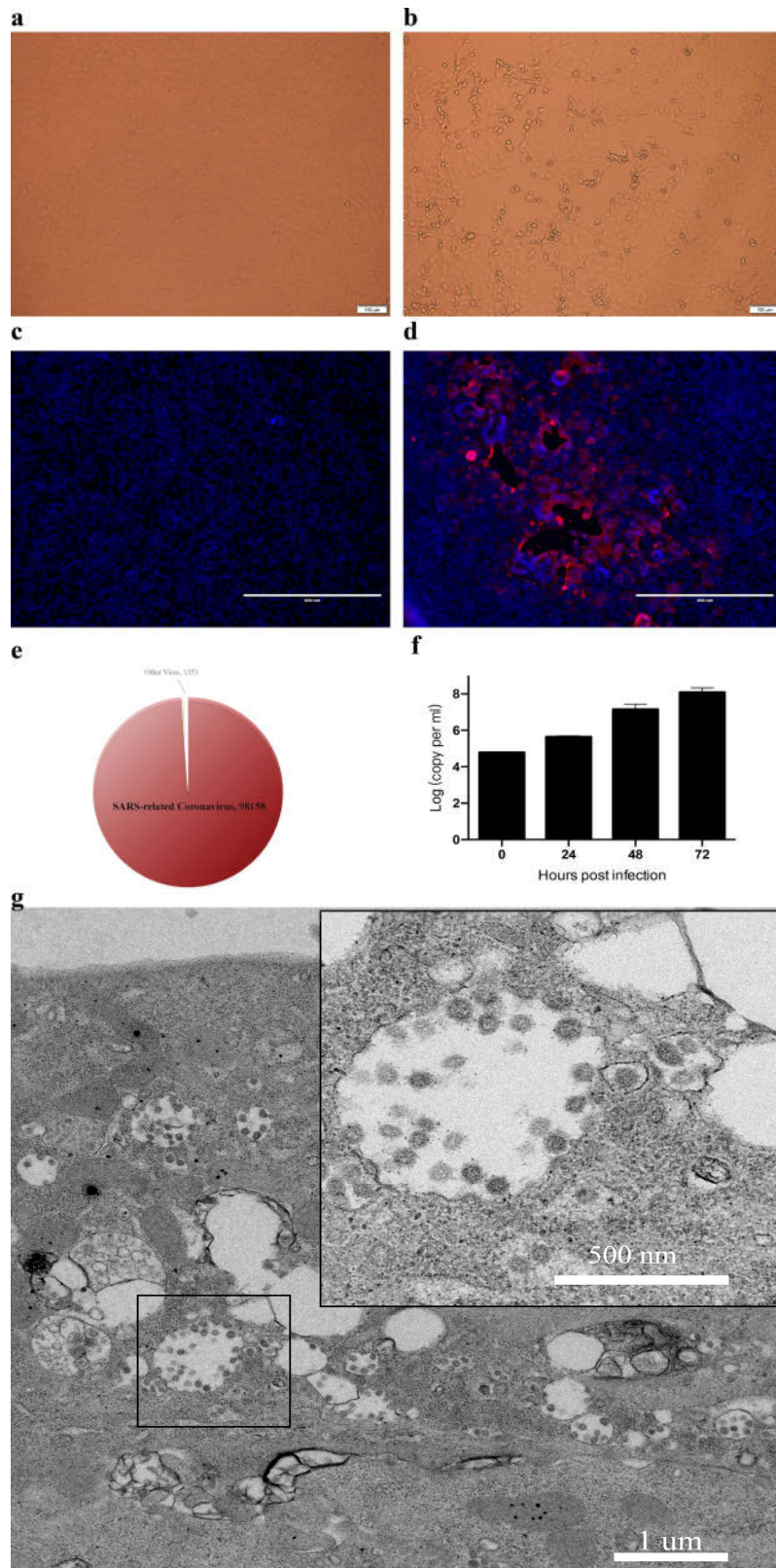


Extended Data Fig. 4 | Molecular detection method used to detect 2019-nCoV. **a**, Standard curve for qPCR primers. The PCR product of the S gene that was serially diluted in the range of 10^8 to 10^1 (lines from left to right) was used

as a template. Primer sequences and experimental conditions are described in the Methods. **b**, Specificity of the qPCR primers. Nucleotide samples from the indicated pathogens were used.

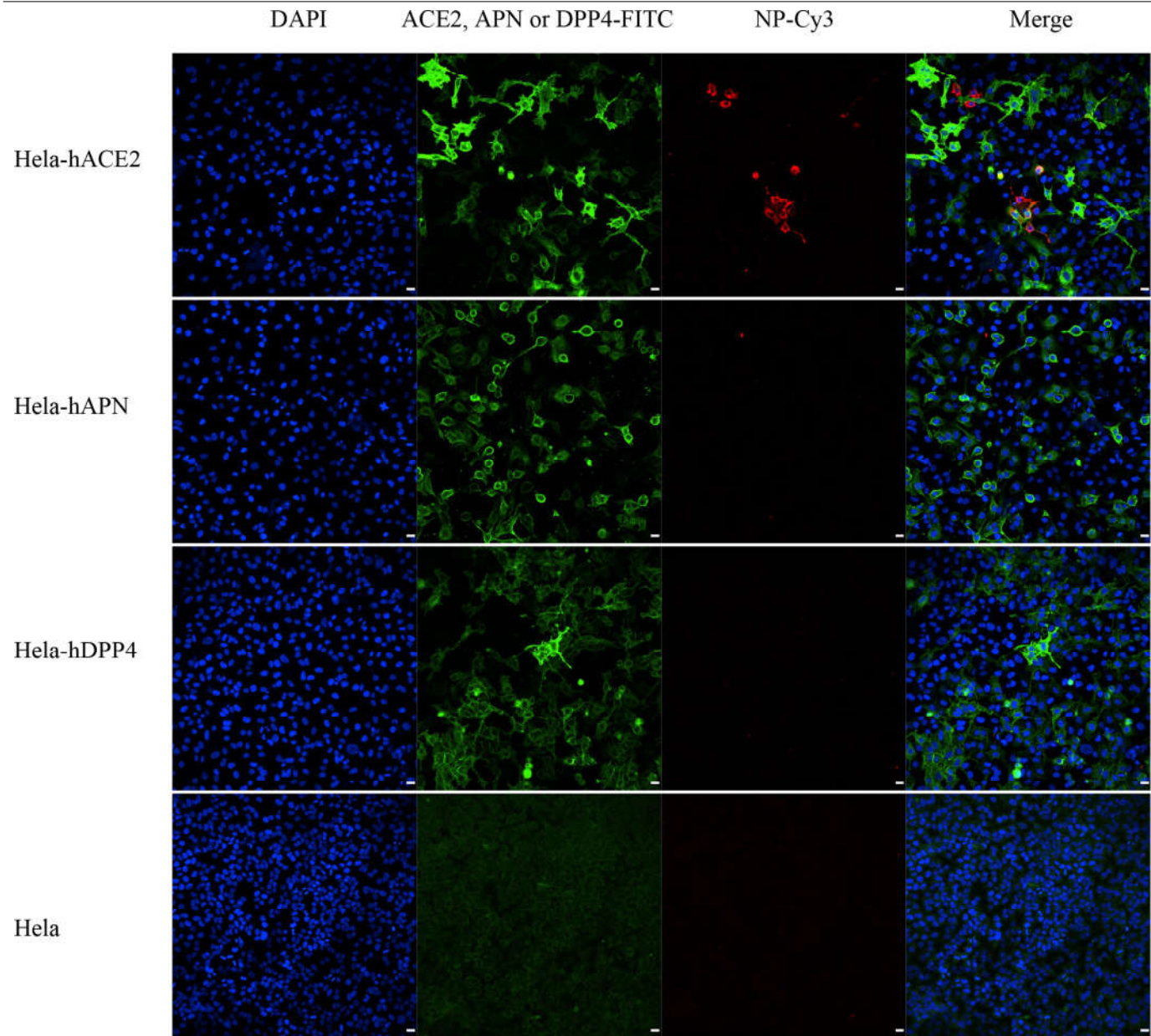


Extended Data Fig. 5 | Amino acid sequence alignment of the nucleocapsid protein of 2019-nCoV to bat SARSr-CoV Rp3 and SARS-CoV BJ01. Bat SARSr-CoV Rp3 was obtained from *R. sinicus*, which is found in Guangxi province.



Extended Data Fig. 6 | Isolation and antigenic characterization of 2019-nCoV. **a, b**, Vero E6 cells are shown at 24 h after infection with mock virus (**a**) or 2019-nCoV (**b**). **c, d**, Mock-virus-infected (**c**) or 2019-nCoV-infected (**d**) samples were stained with rabbit serum raised against recombinant SARSr-CoV Rp3 N protein (red) and DAPI (blue). The experiment was conducted twice independently with similar results. **e**, The ratio of the number of reads related

to 2019-nCoV among the total number of virus-related reads in metagenomics analysis of supernatants from Vero E6 cell cultures. **f**, Virus growth in Vero E6 cells. **g**, Viral particles in the ultrathin sections were imaged using electron microscopy at 200 kV. The sample was from virus-infected Vero E6 cells. The inset shows the viral particles in an intra-cytosolic vacuole.



Extended Data Fig. 7 | Analysis of 2019-nCoV receptor usage. Determination of virus infectivity in HeLa cells with or without the expression of human APN and DPP4. The expression of ACE2, APN and DPP4 plasmids with S tag were detected using mouse anti-S tag monoclonal antibody. ACE2, APN and DPP4 proteins (green), viral protein (red) and nuclei (blue) are shown. Scale bars, 10 μ m.

Extended Data Table 1 | Patient information and diagnosis history

Patient No.	Gender	Age	Date of Onset	Date of Admission	Symptoms When Admitted	Current Status (2020.01.13)	Diagnosis history
ICU-01*	Male	62	2019.12.12	2019.12.27	fever	recover, discharged	negative
ICU-04	Male	32	2019.12.19	2019.12.29	fever, cough, dyspnea	fever, intermittent cough	negative
ICU-05	Male	40	2019.12.17	2019.12.27	fever (38 °C), expectoration, malaise, dyspnea	fever, malaise, intermittent cough	AdV (IgM)
ICU-06	Female	49	2019.12.23	2019.12.27	fever (37.9 °C), palpitation	fever, malaise, cough	Coronavirus (nt)
ICU-08	Female	52	2019.12.22	2019.12.29	fever (38.5 °C), expectoration, malaise, dyspnea	recover, discharged	Streptococcus pneumoniae (nt)
ICU-09	Male	40	2019.12.22	2019.12.28	fever (38.5 °C), expectoration	fever (38.5 °C), malaise, expectoration, dizziness	negative
ICU-10	Male	56	2019.12.20	2019.12.20	fever, dyspnea, chest tightness	fever, malaise, cough, dyspnea	negative

Note, some records are missing. All patients are sellers or deliverymen at the seafood market except ICU-01, whose contact history is unclear. All patients were admitted to intensive care unit (ICU) during the first investigation and were now in stable condition. Blood IgM tests have been performed for the following respiratory pathogens for all patients: *Legionella pneumophila*, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, respiratory syncytial virus, adenovirus, *Rickettsia*, influenza A virus, influenza B virus and parainfluenza virus.

*This patient reported fever on 12 December 2019 and then recovered without medical treatment. He came back to the hospital on 27 December 2019 with a fever. His wife was also ill and admitted to the hospital. Both individuals recovered.

Extended Data Table 2 | Laboratory results

Patient No.	Test No.	First sampling-2019.12.30			Second sampling-2020.01.10			
		BALF	Oral Swab	Blood (Ab)	Oral Swab	Anal Swab	Blood (PCR)	Blood (Ab)
ICU-01	WIV01	-	Ct=32.0	NA	NA	NA	NA	NA
ICU-04	WIV02 [#]	Ct=17.6	Ct=26.6	NA	-	-	-	+
ICU-05	WIV03	Ct=27.0	Ct=31.9	NA	-	-	-	+
ICU-06	WIV04 ^{##}	Ct=18.3	Ct=27.7	+	-	-	-	+
ICU-08	WIV05 [#]	Ct=24.1	-	NA	NA	NA	NA	NA
ICU-09	WIV06 [#]	Ct=21.6	Ct=29.4	NA	-	-	-	+
ICU-10	WIV07 [#]	Ct=25.7	Ct=24.0	NA	-	-	-	+

Samples from two patients (ICU-01 and ICU-08) were not available during the second investigation. They had been discharged from the hospital. We did a serial test for patient ICU-06 on the following dates: 30 December 2019, 31 December 2019, 1 January 2020 and 10 January 2020, corresponding to 7, 8, 9 and 18 days after disease onset (23 December 2019). Molecular and serological (IgM and IgG) virus-detection results for 2019-nCoV are shown. NA, not available.

[#]Virus isolated.

^{##}A full-length genome was obtained.

Extended Data Table 3 | Genomic comparison of 2019-nCoV WIV04 with SARS-CoVs and bat SARSr-CoVs

Sequence identities with SARS-CoVs & bat SARSr-CoVs (nt/aa %)												
	Full-length genome	ORF1a	ORF1b	S	ORF3a	E	M	ORF6	ORF7a	ORF7b	ORF8	N
SARS-CoV GZ02	79.6	76.0/80.9	86.2/95.7	73.4/77.0	75.6/73.4	94.7/96.0	85.4/90.5	76.3/68.9	82.8/86.0	84.8/81.4	52.0/31.6	87.7/91.2
SARS-CoV BJ01	79.6	76.0/80.8	86.2/95.7	73.4/76.9	75.3/72.6	94.7/96.0	85.6/90.5	75.8/67.2	82.8/86.0	84.8/81.4	51.1/-	88.8/91.2
SARS-CoV Tor2	79.6	76.0/80.9	86.2/95.8	73.4/76.7	75.4/72.6	94.7/96.0	85.6/90.5	76.3/68.9	82.8/86.0	84.8/81.4	51.1/-	88.8/91.2
SARS-CoV SZ3	79.6	76.0/81.0	86.2/95.8	73.4/76.9	75.4/72.6	94.7/96.0	85.3/90.0	76.3/68.9	82.8/86.0	84.8/81.4	52.3/31.6	88.8/91.2
SARS-CoV PC4-227	79.5	76.0/80.8	86.1/95.6	73.4/76.7	75.5/72.6	94.7/96.0	85.1/90.0	75.8/68.9	82.8/86.0	84.8/81.4	52.3/-	88.5/90.7
Bat SARSr-CoV RaTG13	96.2	96.0/98.0	97.3/99.3	93.1/97.7	96.3/97.8	99.6/100	95.5/99.6	98.4/100	95.6/97.5	99.2/97.7	97.0/95.0	96.9/99.0
Bat SARSr-CoV WIV1	79.7	76.0/80.7	85.9/95.8	73.4/77.6	76.1/74.5	95.6/96.0	84.8/90.0	78.0/73.8	85.0/88.4	85.6/83.7	65.8/57.9	88.5/90.9
Bat SARSr-CoV WIV16	79.7	75.9/81.0	86.1/95.6	73.1/77.8	76.1/74.5	95.6/96.0	84.8/90.0	77.4/72.1	85.0/88.4	85.6/83.7	65.3/57.9	88.6/90.9
Bat SARSr-CoV SHC014	79.6	75.9/80.9	85.9/95.8	73.3/77.7	76.1/74.5	95.6/96.0	84.8/90.0	78.0/70.5	84.4/88.4	85.6/83.7	65.8/58.7	88.6/90.9
Bat SARSr-CoV Rs4231	79.7	76.0/81.0	86.2/95.8	72.9/77.5	75.8/74.1	94.3/94.7	84.4/90.0	76.9/67.2	85.0/88.4	85.6/83.7	65.3/57.9	88.8/91.4
Bat SARSr-CoV YNLF31C	79.0	75.7/80.6	85.8/95.7	71.4/75.5	75.0/71.2	94.3/96.0	84.7/89.6	76.9/70.5	83.1/87.6	86.4/83.7	50.3/31.3	88.3/90.5
Bat SARSr-CoV LYRa11	79.6	75.8/80.6	85.7/95.6	73.9/77.3	77.2/76.3	94.7/94.7	85.1/90.0	78.5/70.5	82.0/85.1	81.1/81.4	66.7/57.9	89.0/91.6
Bat SARSr-CoV ZC45	88.1	91.0/95.7	86.1/96.0	77.8/82.3	87.8/90.9	98.7/100	93.4/98.6	95.2/93.4	88.8/87.6	94.7/93.0	88.5/94.2	91.1/94.3
Bat SARSr-CoV ZXC21	88.0	90.9/95.7	86.2/95.8	77.1/81.7	88.9/92.0	98.7/100	93.4/98.6	95.2/93.4	89.1/88.4	95.5/93.0	88.5/94.2	91.2/94.3
Bat SARSr-CoV HuB2013	79.6	76.3/81.2	85.3/95.7	73.1/76.8	75.4/75.5	95.2/94.7	85.3/91.0	76.3/68.9	84.2/87.6	85.6/83.7	62.0/49.6	88.9/91.6
Bat SARSr-CoV GX2013	79.1	75.9/80.8	86.0/95.9	73.1/77.1	75.6/73.0	94.7/96.0	84.8/91.4	77.4/68.9	85.0/86.8	84.1/79.1	51.4/31.6	87.9/90.2
Bat SARSr-CoV SX2013	78.9	76.2/80.6	85.1/95.5	71.2/75.5	74.7/71.2	94.3/93.3	83.0/89.6	77.4/68.9	84.2/86.8	85.6/83.7	49.7/30.4	86.9/90.2
Bat SARSr-CoV SC2018	79.4	75.8/80.7	85.5/95.2	72.7/76.4	75.0/71.2	94.3/96.0	84.7/90.0	80.0/71.8	85.2/87.6	84.8/83.7	66.1/55.4	88.2/91.2
Bat SARSr-CoV Rs672	79.6	76.0/80.9	85.9/95.8	72.8/76.2	75.2/71.9	95.2/96.0	84.8/89.6	78.5/70.5	84.7/88.4	85.6/83.7	65.8/58.7	87.9/91.2
Bat SARSr-CoV Rp3	79.5	75.9/80.5	86.0/95.7	73.1/77.2	74.9/74.8	95.2/96.0	85.1/90.0	76.9/68.9	83.9/89.3	84.8/83.7	66.4/56.2	88.4/90.7
Bat SARSr-CoV Rf1	78.8	76.2/80.6	84.8/95.3	71.1/75.7	74.3/69.0	94.3/94.7	83.3/89.6	79.0/68.9	84.2/86.8	84.1/83.7	50.6/31.3	86.8/89.5
Bat SARSr-CoV HKU3-1	79.4	76.1/80.9	84.9/95.1	73.4/77.9	75.8/73.4	95.2/96.0	84.7/91.0	75.3/67.2	85.0/89.3	84.1/79.1	66.4/57.0	88.3/90.0

Extended Data Table 4 | Virus neutralization test of serum samples

Samples	VNT titre for nCoV-2019
Healthy people #1 from Wuhan	neg
Healthy people #2 from Wuhan	neg
Horse anti-SARS-CoV serum	>1:80
WIV02	>1:80
WIV03	1:40
WIV04	>1:80
WIV06	>1:80
WIV07	>1:80

Each serum sample was tested in triplicate. Serum samples from two healthy individuals from Wuhan and five patients as well as a horse anti-SARS-CoV anti-serum were used. We used 120 TCID₅₀ viruses per well. Serum samples were used at dilutions of 1:10, 1:20, 1:40 and 1:80. neg, negative; VNT, virus neutralization test.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used.

Data analysis

BWA (v0.7.12-r1039), Cutadapt (v1.18), Geneious (v11.0.3), MEGAHIT (v1.2.9), Clone Manager Professional Suite 8, MAFFT (v7.307), MGmapper (PE2.24 and SE2.24), PAL2NAL (version 14), Clustal Omega (version 1.2.4), RAxML (version 0.9.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that support the findings of this study have been deposited in GISAID with the accession no. EPI_ISL_402124 and EPI_ISL_402127-402131.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples of seven pneumonia patients are available from the clinical hospital to be sent to Wuhan Institute of Virology for pathogen identification. The coronavirus genome sequences were obtained from 5 different patients and shared >99.9% identity, suggesting they were infected by the same virus. Therefore, the sample size is sufficient for conducting the following study which aims to identify and characterize the causative agent of this pneumonia outbreak.
Data exclusions	No data excluded
Replication	The authors guarantee the findings are reliably reproducible. At least three independent experiments were performed, which was stated in the text.
Randomization	Samples were chosen randomly.
Blinding	We were blinded when choosing samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	1. SARSr-CoV Rp3 NP antibody made in house; 2. Anti-Human IgG-HRP conjugated monoclonal antibody (Kyab Biotech Co., Ltd, Wuhan, China, dilution: 1:40000); 3. Anti-Rp3 NP-HRP conjugated (Kyab Biotech Co., Ltd, Wuhan, China, dilution: 1:4000); 4. FITC-labelled goat anti-mouse IgG H&L (Abcam, ab96879, dilution 1:100); 5. cyanin 3-conjugated goat anti-rabbit IgG (Abcam, ab6939, dilution: 1:200); 6. mouse anti-S tag monoclonal antibody made in house
Validation	The house-made SARSr-CoV Rp3 NP antibodies and anti-S tag monoclonal antibody were validated in a WB. The cy3-conjugated anti-rabbit IgGs were validated in IFA. The FITC-labelled goat anti-mouse IgG H&L was validated in IHC.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	1. African green monkey origin, Vero E6 cell; 2. Human lung cell Huh7 ; 3. Human HeLa cells. All cell lines were from ATCC.
Authentication	All monkey and human cells were from ATCC with authentication. The authentication was performed by microscope morphology check, growth curve analysis or identity verification with STR analysis (for human cell lines).
Mycoplasma contamination	We confirm that all cells were tested as mycoplasma negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Participants were all 2019-nCoV infected patients.
Recruitment	Samples were sent to Wuhan Institute of Virology by hospital for pathogen identification.
Ethics oversight	Wuhan Jinyintan Hospital (the co-authored institution)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Peripheral T cell expansion predicts tumour infiltration and clinical response

<https://doi.org/10.1038/s41586-020-2056-8>

Received: 27 September 2018

Accepted: 8 January 2020

Published online: 26 February 2020

 Check for updates

Thomas D. Wu^{1✉}, Shravan Madireddi², Patricia E. de Almeida², Romain Banchereau³, Ying-Jiun J. Chen⁴, Avantika S. Chitre², Eugene Y. Chiang², Hina Iftikhar², William E. O’Gorman⁵, Amelia Au-Yeung⁵, Chikara Takahashi⁵, Leonard D. Goldstein¹, Chungkee Poon⁶, Shilpa Keerthivasan², Denise E. de Almeida Nagata², Xiangnan Du², Hyang-Mi Lee², Karl L. Banta², Sanjeev Mariathasan³, Meghna Das Thakur⁷, Mahrukh A. Huseni⁷, Marcus Ballinger⁷, Ivette Estay⁷, Patrick Caplazi⁸, Zora Modrusan⁴, Lélia Delamarre², Ira Mellman², Richard Bourgon¹ & Jane L. Grogan^{2,9✉}

Despite the resounding clinical success in cancer treatment of antibodies that block the interaction of PD1 with its ligand PDL1¹, the mechanisms involved remain unknown. A major limitation to understanding the origin and fate of T cells in tumour immunity is the lack of quantitative information on the distribution of individual clonotypes of T cells in patients with cancer. Here, by performing deep single-cell sequencing of RNA and T cell receptors in patients with different types of cancer, we survey the profiles of various populations of T cells and T cell receptors in tumours, normal adjacent tissue, and peripheral blood. We find clear evidence of clonotypic expansion of effector-like T cells not only within the tumour but also in normal adjacent tissue. Patients with gene signatures of such clonotypic expansion respond best to anti-PDL1 therapy. Notably, expanded clonotypes found in the tumour and normal adjacent tissue can also typically be detected in peripheral blood, which suggests a convenient approach to patient identification. Analyses of our data together with several external datasets suggest that intratumoural T cells, especially in responsive patients, are replenished with fresh, non-exhausted replacement cells from sites outside the tumour, suggesting continued activity of the cancer immunity cycle in these patients, the acceleration of which may be associated with clinical response.

We sequenced 330 million mRNA transcripts in 141,623 T cells from 14 treatment-naïve patients across four different types of cancer (Extended Data Fig. 1a–i, Supplementary Tables 1, 2). For each patient, samples were obtained from surgically resected tumour and histologically normal adjacent tissue (NAT), and, in four patients, from peripheral blood. Single-cell T cell receptor (TCR) sequencing (scTCR-seq) yielded one or more complementarity-determining region 3 (CDR3) regions of α - or β -chains in 99,788 cells. T cells were grouped into 56,975 distinct clonotypes by matching CDR3 regions, which allowed us to measure clonal expansion and track clonal lineages across tissues. Although individual clonotypes were rarely shared across patients (Supplementary Table 3), many clonotypes resided in both the tumour and the corresponding NAT from a given patient (Extended Data Fig. 2a). Such dual-expanded clones contrasted with expanded clones that resided solely in one compartment (termed tumour or NAT multipliers), and with clones having only one observed cell in either compartment (singletons).

Although most clonotypes were singletons (Extended Data Fig. 2b), representing a diverse repertoire of T cells, 9–18% of clonotypes in each

patient were clonally expanded multipliers or dual-expanded clones. Dual-expanded clonotypes constituted substantial fractions of observed cells, sometimes the majority of T cells in a given compartment (Extended Data Fig. 2c). Analysing clones by their cell counts in NAT and tumour samples revealed diverse clonal expansion behaviours across patients (Extended Data Fig. 2d). In some patients—for example, Colon1—dual-expanded clones aligned along the main diagonal, reflecting approximately equal cell counts in the two compartments and suggesting coordinate or parallel processes in tumour and NAT, brought about by multiple clonotypes acting similarly. Conversely, in other patients, dual-expanded clones were scattered from the main diagonal, indicating that each clone had an independent, divergent pattern of expansion, migration or attrition in the tumour and NAT compartments.

We hypothesized that parallel expansion in tumour and NAT might represent infiltration of T cells from the periphery, extravasating into both tissues equally from inflamed blood vessels². We found supporting evidence from matched blood samples, in which clones highly expanded in blood showed evidence of parallel expansion in

¹Department of Bioinformatics and Computational Biology, Genentech, Inc., South San Francisco, CA, USA. ²Department of Cancer Immunology, Genentech, Inc., South San Francisco, CA, USA. ³Department of Oncology Biomarker Development, Genentech, Inc., South San Francisco, CA, USA. ⁴Department of Microchemistry, Proteomics, Lipidomics and Next Generation Sequencing, Genentech, Inc., South San Francisco, CA, USA. ⁵Department of OMNI Biomarker Development, Genentech, Inc., South San Francisco, CA, USA. ⁶Department of Research Biology, Genentech, Inc., South San Francisco, CA, USA. ⁷Department of Development Sciences, Genentech, Inc., South San Francisco, CA, USA. ⁸Department of Pathology, Genentech, Inc., South San Francisco, CA, USA. ⁹Present address: ArsenalBio, South San Francisco, CA, USA. ✉e-mail: twu@gene.com; jane@arsenalbio.com

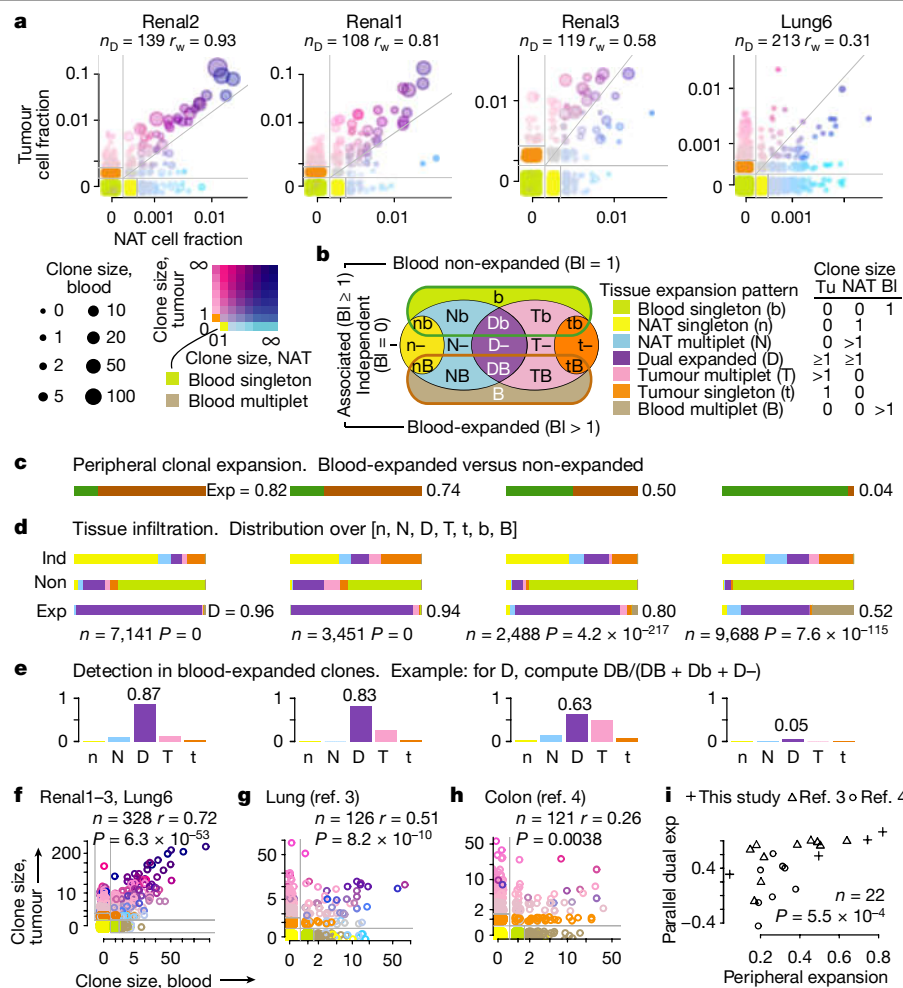


Fig. 1 | Parallel dual expansion and peripheral clonal expansion. **a**, Clonal expansion scatter plots. Scatter plots of clonotypes are shown for patients with blood samples, plotted by normalized clone size, or cell fraction, in NAT and tumour, with random jitter to distinguish points. Dots are sized for blood clone size, and coloured by a palette for tissue expansion pattern. Diagonal lines indicate equal cell fractions. Other lines separate the absence and presence of clones within compartments. r_w denotes Pearson's correlation coefficient (r), weighted by $(1 + \text{blood clone size})$, on the dual-expanded clones (n_D). **b**, Schema for clonal expansion. Tissue and blood expansion patterns are defined by clones sizes in tumour (Tu), NAT and blood (BI). **c**, Peripheral clonal expansion. Bar plots for each patient in **a** show fractions of cells from blood-expanded (values shown) versus non-expanded clones, coloured as in **b**, **d**, Tissue infiltration. Bar plots show distributions of cells by tissue expansion patterns from blood-independent (Ind), non-expanded (Non) and expanded clones (Exp). Values for dual-expanded clones (D) are shown. P values were determined by a chi-square test on counts of tissue-resident cells (those in tumour or NAT but not in blood). **e**, Detection of tissue TCRs in blood. Bar plots show fractions of tissue-resident cells with clonotypes observed in a blood-expanded clone, for each tissue expansion pattern (values for D are shown). n, NAT singleton; N, NAT multiplet; t, tumour singleton; T, tumour multiplet. **f–h**, Correlation of blood and tumour clone sizes. Scatter plots show blood and tumour clone sizes from patients in **a** and external datasets^{3,4}, coloured as in **a**. Horizontal and vertical lines separate clone sizes of zero and non-zero. Two-sided P values are shown from a Pearson correlation coefficient r on log-transformed clone sizes for clones present in both tumour and blood. **i**, Peripheral expansion and parallel dual expansion. Values are summarized for patients in **f–h**. P values were determined by a two-tailed linear regression t -test, after exponentiating y values to linearize the data, across patients.

tumour and NAT (Fig. 1a). To quantify our observations in these four patients, we classified clones not only by their expansion in tissue but also by their peripheral expansion, as measured by clone sizes in blood (Fig. 1b, Extended Data Fig. 2e). Clones with single or multiple constituent cells in blood were termed blood non-expanded and blood-expanded, respectively, whereas clones without any associated cells in blood were termed blood-independent. With this schema, we found that an aggregate measure of peripheral expansion over all clones—as determined by the fraction of blood T cells from blood-expanded versus non-expanded clones (Fig. 1c)—associated with higher levels of lymphocytic infiltration by blood-expanded clones into tumour and NAT samples, and served as a proxy for the extent of dual expansion (Fig. 1d). Notably, in patients with strong dual expansion, dual-expanded clones in tissue could frequently be detected as expanded in blood (Fig. 1e).

Our scTCR-seq data allowed us to make observations at the level of clones rather than individual T cells. By studying clones, we found that peripheral and intratumoural clone sizes were significantly correlated ($P = 6.3 \times 10^{-53}$, $P = 8.2 \times 10^{-10}$ and $P = 0.0038$) in our dataset as well as in two external datasets^{3,4} (Fig. 1f–h), which indicates that the relationship between peripheral expansion and tumour infiltration held not only for aggregate cell fractions but also for individual clones. More notably, a general and significant ($P = 5.5 \times 10^{-4}$) association was observed across patients in the combined datasets between the total amount of peripheral expansion and parallel dual expansion (Fig. 1i, Extended Data Fig. 3). Collectively, our results show a strong

relationship between peripheral clonal expansion and parallel infiltration into tumours and NAT.

We explored the phenomenon of clonal expansion further by simultaneously measuring gene expression using single-cell RNA sequencing (scRNA-seq). Transcriptional profiles of individual T cells allowed grouping of similar cells into clusters (Fig. 2a), which we characterized by cross-labelling our cells against reference gene signatures from published datasets^{3–5} (Extended Data Fig. 4a), and by profiling genes we found to be cluster-specific (Extended Data Fig. 4b, Supplementary Table 4). Cluster analysis identified T effector cells (8.1- T_{eff}), T effector memory cells (8.2- T_{em}), and regulatory T cells (4.6a- T_{reg} , 4.6b- T_{reg}), as well as subtypes or physiological states of resident memory T (T_{rm}) cells⁶, which shared abundant expression of *ITGAE* (also known as *CD103*) and *ZNF683* (*Hobit*), a regulator of tissue residency⁷, as well as positive scores on a published gene signature of T_{rm} cells⁸ (Extended Data Fig. 4c). These subclusters, designated 8.3a- T_{rm} , 8.3b- T_{rm} and 8.3c- T_{rm} , differed by low, intermediate and high expression of activation-exhaustion markers, respectively, including *PDI* (also known as *PDCD1*) (Extended Data Fig. 4d), and also by a published gene signature for terminally exhausted versus stem-like exhausted $CD8^+$ T cells⁹ (Extended Data Fig. 4e), consistent with recently identified phenotypes of progenitor exhausted and terminally exhausted T_{rm} cells¹⁰. Subcluster 8.3c had lower expression of activation marker *XCL1* (which encodes lymphotactin), while 8.3b had exclusively high expression of *KLRC2* (also known as *NKG2-C*) and *KLRC3* (*NKG2-E*) (Extended Data Fig. 4b), observed in $CD8^+$ T cells with both TCR-dependent and innate-like,

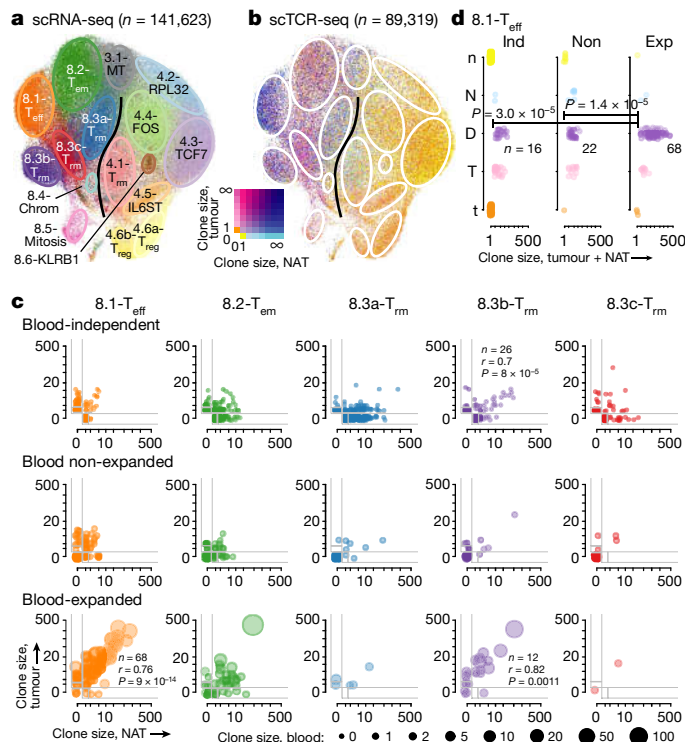


Fig. 2 | T cell clusters and clonal expansion. **a**, Cluster analysis. T cells from all 14 patients are plotted by uniform manifold approximation and projection (UMAP) dimensionality reduction of the scRNA-seq data, and coloured by cluster assignment, with cluster regions approximated by labelled ovals. The curved line highlights the separation of CD8 (left) and CD4 (right) T cells. **b**, Tissue expansion patterns by cluster. T cells in tumour and NAT with a clonotype are plotted by UMAP coordinates from **a** and coloured by tissue expansion pattern of their parent clone using a two-dimensional palette. **c**, Clonal expansion by cluster and blood expansion pattern. Clones from Fig. 1a are plotted by blood independence, non-expansion and expansion (rows) and by primary cluster (columns), using coordinates for NAT and tumour clone size, and dot size for blood clone size. Grey lines separate NAT and tumour clone sizes of zero and non-zero. Two-sided P values are shown when $P < 0.01$ from a Pearson's correlation coefficient r on log-transformed clone sizes for dual-expanded clones. **d**, Analysis of 8.1- T_{eff} clones. Scatter plots from **c** for 8.1- T_{eff} clones are shown as distributions of clone sizes in tumour plus NAT for each tissue expansion pattern, abbreviated as in Fig. 1b. Vertical jitter distinguishes overlapping points. Two-sided P values are shown for Bonferroni-adjusted $P < 0.01$ from a t -test on log-transformed clone sizes for dual-expanded clones.

TCR-independent mechanisms of activation¹¹. In addition, subclusters 8.3a to 8.3c showed increasing prevalence in tumour versus NAT (Extended Data Fig. 4f), supporting a relationship between *PDI* expression, terminal exhaustion, and increased exposure to cognate antigens within the tumour microenvironment¹².

In addition, our deep sequencing analysis revealed several clusters of T cells not matching published reference gene signatures (Extended Data Fig. 4a). We characterized these clusters using gene set enrichment analysis, which showed active histone modification and mitosis in clusters 8.4 and 8.5, respectively (Extended Data Fig. 4g). Chromatin regulation in 8.4 was further evidenced by high expression of the long noncoding RNA *MALAT1*¹³ and the chromatin remodelling enzyme *CHD1*¹⁴ (Extended Data Fig. 4b). Cluster 3.1 expressed abundant mitochondrial genes, suggestive of apoptosis, and we set this cluster aside as generic CD3⁺ cells. Clusters 8.4-Chrom and 8.5-Mitosis had similarities to T_{rm} cells, with cells expressing a range of *ITGAE* levels (Extended Data Fig. 4b) and the highest mean expression of *PDI* (Extended Data Fig. 4d), generally considered a marker of exhaustion. Moreover, 8.5-Mitosis

also cross-labelled most closely with clusters that have been reported to represent exhausted T cells, specifically, the CD8_{ex} cluster observed by Yost et al.⁵ and the CD8_{C6-LAYN} cluster observed by Guo et al.³ (Extended Data Fig. 4a). However, neither 8.4-Chrom nor 8.5-Mitosis scored highly on the published signature for T_{rm} cells, which are thought to be the main source of exhausted T cells (Extended Data Fig. 4c).

Combining scTCR-seq and scRNA-seq data yielded insights into the clonal expansion behaviour of clones and T cells. One challenge in such combined analysis is that whereas expansion behaviour is attributable to clones, clusters are assigned to individual cells based on their transcriptional profiles. Therefore, a given clone may comprise a mixture of T cell phenotypes with distinct transcriptional profiles and hence different clusters. Accordingly, clones in our dataset exhibited a range of transcriptional diversity: some clones were homogeneous for a single cluster, whereas others had heterogeneous cellular mixtures of several clusters (Extended Data Fig. 4h, i). Nevertheless, most clones had a primary cluster that predominated and served as an approximate phenotype for subsequent analysis.

As expected, clones partitioned into CD8 or CD4 types, which exhibited distinct tissue expansion behaviours (Extended Data Fig. 5a, b). Clones of primarily 8.1, 8.2 and 8.3 T cells were largely dual-expanded (Fig. 2b), whereas clones of CD4⁺ T cells were generally singletons, with some exceptions: (1) low-level dual expansion by 4.4-FOS clones; (2) tumour and NAT multiplet expansion by clones of 4.1- T_{rm} cells; and (3) tumour multiplet expansion by 4.5-IL6ST and 4.6- T_{reg} clones (Extended Data Fig. 5c). Clusters 8.4, 8.5 and 8.6 showed very little clonal expansion in tissue. Our analysis of external datasets^{3,4} yielded similar results, with the analogous CX3CR1 and GZMK clusters showing similar patterns of clonal expansion to our 8.1- T_{eff} and 8.2- T_{em} clusters, respectively (Extended Data Fig. 5d, e).

When we mapped primary clusters onto clones from our four patients with blood samples (shown previously in Fig. 1a), we observed that many clones exhibiting parallel dual expansion were 8.1- T_{eff} clones (Extended Data Fig. 6a). To investigate this association, we re-grouped clones from the patients according to their blood expansion patterns and primary cluster (Fig. 2c), and found that blood-expanded clones with a primary cluster of 8.1- T_{eff} had a significant ($P = 9 \times 10^{-14}$) correlation between tumour and NAT clone sizes. In addition, 8.3b- T_{rm} clones, possibly containing T cells with TCR-independent activation, showed lower levels of parallel dual expansion. Nevertheless, 8.1- T_{eff} clones were more numerous and were the only ones to show a statistically significant ($P = 1.4 \times 10^{-5}$ and $P = 3.0 \times 10^{-5}$) relationship between blood expansion and dual-expanded clone size (Fig. 2d, Extended Data Fig. 6b), indicating their predominant role in peripheral clonal expansion and tumour infiltration.

Blood-expanded 8.1- T_{eff} clones may have a role not only in tumour infiltration but also in differentiation into other T cell phenotypes¹⁵. Although blood-expanded 8.1- T_{eff} clones constituted only a small fraction of all clones (Extended Data Fig. 6c, d), they had a disproportionate influence on T cell composition in blood and tumours. By linking individual T cell phenotypes with the blood expansion patterns of their parent clones, we observed that 8.2- T_{em} and 8.3- T_{rm} cells in blood (Extended Data Fig. 6e, f) and tumours (Extended Data Fig. 6g, h) shared clonotypes extensively with blood-expanded clonal lineages of 8.1- T_{eff} cells. Thus, differentiation of blood-expanded 8.1- T_{eff} clones may yield a diverse array of T cell subtypes in blood and tumours. By contrast, most 8.2- T_{em} and 8.3- T_{rm} cells in NAT belonged to blood-independent clones (Extended Data Fig. 6i, j), consistent with mechanisms of local expansion.

Our data capture only a single snapshot of clone sizes across tumour, NAT and blood, representing the accumulation over time of various dynamic processes. To understand these dynamics, we analysed an external dataset⁵ of pre- and post-treatment T cells from basal cell carcinomas, with the understanding that the two sets of measurement differ both by the effects of immunotherapy and by the passage of time.

The published analysis demonstrated that new clones appear after treatment, with clonotypes distinct from those detected before treatment. We extended the analysis by matching novel clones in tumours after treatment with pre-existing clones in blood before treatment, as assayed by bulk TCR sequencing (TCR-seq) (Extended Data Fig. 7a). We found significant ($P = 2.3 \times 10^{-8}$ and $P = 1.3 \times 10^{-30}$) correlations in clone size and transcript counts among novel CD8⁺ cells (Fig. 3a, Extended Data Fig. 7b) but not novel CD4⁺ cells (Extended Data Fig. 7c), which suggests that at least some of the new CD8⁺ clones originated peripherally. Moreover, although patients differed substantially in number and expansion of new clones, these differences associated with the degree of clonal diversity in peripheral blood (Fig. 3b, Extended Data Fig. 7d), further supporting a relationship between peripheral and intratumoural clonal expansion.

We further categorized new tumour clones based on whether they shared TCR sequences with blood before treatment, with the caveat that incomplete sampling of TCRs in blood (Extended Data Fig. 7e) may misidentify some clones as being blood-independent. Because bulk TCR-seq measures total TCR transcripts rather than cells with a given TCR, we could not discriminate blood-expanded from non-expanded clones, but we could distinguish blood-independent from blood-associated clones by the absence or presence of a matching bulk TCR-seq transcript. Notably, in the two patients with greatest peripheral clonal diversity, we observed a statistically significant ($P = 3.8 \times 10^{-16}$ and $P = 2.2 \times 10^{-11}$) interaction between blood association and exhaustion phenotype (Fig. 3c): non-exhausted clones were more likely to be blood-associated, whereas exhausted clones were more likely to be blood-independent. Furthermore, non-exhausted clones that associated with blood had significantly ($P = 3.1 \times 10^{-5}$, $P = 5.2 \times 10^{-12}$ and $P = 2.0 \times 10^{-14}$) greater clone sizes in tumours after treatment than those independent of blood. Although patients available for study were limited, our observations suggest two possible modes of clonal expansion: (1) a local expansion independent of blood (or possibly stemming from peripheral events in the remote past) that generates exhausted T cells; and (2) an infiltrative process from contemporary or relatively recent blood content that introduces novel, non-exhausted clones into the tumour.

The prevalence of infiltrating T cells, which may be reactive not only to the tumour but also to non-tumour antigens, could account for recent observations of 'bystander' T cells that are reactive against non-tumour antigens¹⁶ and might help explain recent findings that only a small fraction of intratumoural CD8⁺ T cells are reactive against tumour¹⁷. We compared our CDR3 amino acid sequences against databases of CDR3 sequences that are known or likely to react to common viral antigens. Although TCRs often matched those corresponding to common viral antigens, almost all of these TCRs belonged to singleton clones and therefore showed no evidence of clonal expansion. A few TCRs in each patient did exhibit high levels of clonal expansion (Extended Data Fig. 8a), which suggests that although bystander T cells may be prevalent at the cellular level, relatively few clones may be responding to non-tumour antigens. We observed over-representation of virally matching TCRs among dual-expanded clonotypes ($P = 3.7 \times 10^{-5}$ and $P = 0.0018$) (Extended Data Fig. 8b) and various CD8 clusters (Extended Data Fig. 8c), which is consistent with the infiltration of clonally expanded bystander T cells from the periphery.

As our analysis indicates, patients vary in their extent of peripheral clonal expansion and resulting infiltration of T cells, and this variability could potentially explain differing clinical responses to immune checkpoint blockade. We evaluated bulk RNA-seq tumour samples from three randomized phase II trials of the anti-PDL1 antibody atezolizumab^{18–20} using gene signatures of tissue expansion patterns involving tumour tissue: tumour singletons, tumour multipliers, and dual-expanded clones (Extended Data Fig. 9a, b, Supplementary Table 5). Because clonal expansion occurs primarily with CD8⁺ T cells, these gene signatures correlated highly with *CD8A* expression (Extended Data Fig. 9c)—an

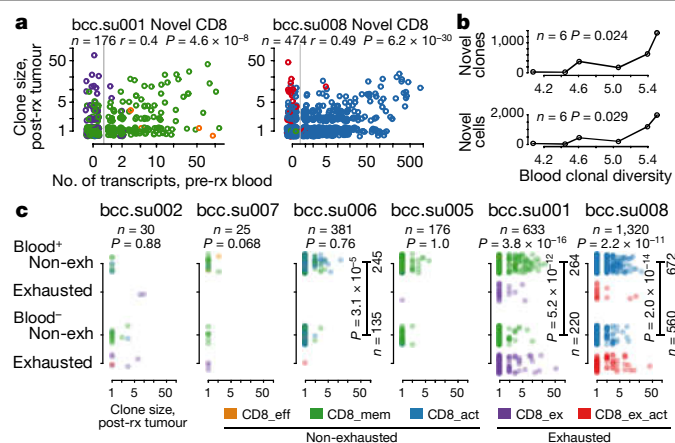


Fig. 3 | Peripheral clonal expansion and novel intratumoural clones.

a, Correlated clone sizes in tumour and blood. Scatter plots of tumour clone size after treatment (post-rx) and transcript counts in blood before treatment (pre-rx) are shown for the two patients in Yost et al.⁵ with bulk TCR-seq blood samples and the largest numbers of novel clones. Vertical lines separate blood-independent (left) and blood-associated (right) clones. Clones are coloured by primary cluster from the original analysis, as in c. Two-sided P values are shown from Pearson's correlation coefficients r on log-transformed values from novel blood-associated CD8 clones. **b**, Blood clonal diversity and novel clones. Data from Extended Data Fig. 7b, d are summarized, showing the total numbers of novel CD8 clones (top) and cells (bottom) after treatment across patients as a function of blood clonal diversity (average of pre- and post-treatment measurements). Two-sided P values are shown from a linear regression t -test across patients, on the logarithm of y values to linearize the data. **c**, Non-exhausted new clones and blood association. Novel clones in Yost et al.⁵ are separated by pre-treatment blood association (Blood⁺) or blood independence (Blood⁻), and further divided by exhaustion or non-exhaustion of their post-treatment primary cluster, as defined and coloured in the legend. Patients are ordered by increasing blood clonal diversity. P values (top) are from a chi-square test on counts of new CD8 clones. P values (side) are from a two-sided t -test on log-transformed clone sizes (shown if $P < 0.01$).

indicator of intratumoural prevalence of CD8⁺ T cells and a reported predictor of response to immunotherapy²¹ (Extended Data Fig. 9d). Nevertheless, the potential exists for improving on *CD8A* expression as a predictor, exemplified by the slightly stronger association of progression-free survival (PFS) with the expression of *CCL5* (also known as *RANTES*)—a marker of T cell activation²² that ranked highly in both tumour multiplet and dual-expansion signatures—in treatment arms in which atezolizumab is active (with monotherapy in the IMmotion150 trial thought to have been mitigated by immunosuppressive myeloid cells²⁰) (Extended Data Fig. 9e).

Similarly, gene signatures for tumour multipliers and dual expansion were associated with greater PFS in relevant arms, both by the hazard ratios of individual genes (Fig. 4a) and by the survival analysis of combined signature scores ($P \leq 0.02$) (Extended Data Fig. 10a). To handle confounding expression of *CD8A*, we performed survival analyses of *CD8A* expression and expansion signature scores as dual variates (Fig. 4b, Extended Data Fig. 10b). Although most patients were concordant in both measures, the relatively few discordant patients with high *CD8A* expression and low expansion signature scores were anomalous and therefore informative. In every atezolizumab-containing arm, the discordant group had worse PFS than the concordant group, with a hazard ratio > 1 in several cases. Furthermore, a similar dual-variate analysis of tumour multiplet and dual expansion signatures showed additive predictive power (Extended Data Fig. 10c), which suggests that these two types of clonal expansion may contribute independently in the clinical response of patients to cancer immunotherapy.

In summary, dual-expanded T cell clones in tumours, and their derivation from peripherally expanded clones, may represent the overall

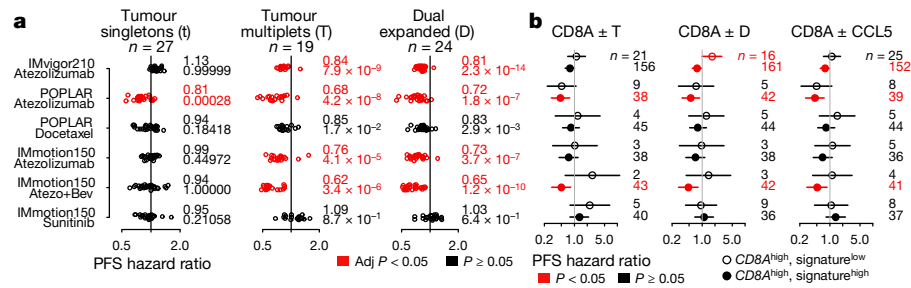


Fig. 4 | Survival analysis of tissue expansion gene signatures. a, Gene hazard ratios. Each gene signature for a tissue expansion pattern involving tumour tissue (columns) is represented by hazard ratios in treatment arms (rows) after dichotomization within the clinical trial. Vertical jitter distinguishes overlapping dots. Hazard ratio < 1 (left of vertical lines) indicates greater PFS. Mean hazard ratio and one-sided P values are shown from a one-sample z -test on hazard ratios for genes in the signature, highlighted with associated data in red when Bonferroni-adjusted $P < 0.05$. **b**, Survival analysis of signatures with

high $CD8A$ expression. Plots summarize results from a dual-variate survival analysis of dichotomized values of $CD8A$ gene expression and tissue expansion signature scores. Points and segments show PFS hazard ratio and 95% confidence intervals from two groups: ' $CD8A^{\text{high}}$, signature^{low}' or ' $CD8A^{\text{high}}$, signature^{high}'. Pairs of groups correspond to each treatment arm in **a**. Vertical lines mark hazard ratio = 1. The number of patients in each group is shown, highlighted with the associated hazard ratio and confidence interval in red when a two-sided $P < 0.05$ from a Cox proportional-hazards model.

strength of immune response, or the 'set point'²³ of the immune system against tumour or foreign antigens. Our study presents an alternative to the widely presumed mechanism that immune checkpoint blockade acts on chronically stimulated T cells to reverse a terminally differentiated or 'exhausted' state observed originally during chronic virus infection^{24,25}. Recent observations have also questioned this model, showing that the phenotype of terminal exhaustion is epigenetically locked and difficult to alter^{26,27}. In addition, recent studies have linked clinical response to stem-like memory $CD8^+$ T cells rather than exhausted ones²⁸. Overall, our study suggests that non-exhausted T cells and T cell clones supplied from the periphery may be key factors in explaining patient variability and clinical benefit from cancer immunotherapy.

Clinical benefit could arise from a direct effect of PD1 or PDL1 blockade on effector T cells²⁹ or other non-exhausted T cells, or because blockade can increase the production or effectiveness of committed anti-tumoural T cells only in those patients with an ongoing T cell response and continued replenishment of tumour-infiltrating lymphocytes. As a practical consequence of this study, the observed correlations between TCR repertoires of dual-expanded clones in tumours and those of peripherally expanded clones suggest that sampling and identifying expanded clones in blood may characterize the TCR composition of clinically relevant intratumoural T cells, expanding the possibilities for 'liquid biopsies'^{30,31}.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2056-8>.

- Mellman, I., Coukos, G. & Dranoff, G. Cancer immunotherapy comes of age. *Nature* **480**, 480–489 (2011).
- Shulman, Z. et al. Transendothelial migration of lymphocytes mediated by intraendothelial vesicle stores rather than by extracellular chemokine depots. *Nat. Immunol.* **13**, 67–76 (2012).
- Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
- Zhang, L. et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
- Yost, K. E. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat. Med.* **25**, 1251–1259 (2019).
- Schenkel, J. M. & Masopust, D. Tissue-resident memory T cells. *Immunity* **41**, 886–897 (2014).
- Mackay, L. K. et al. Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463 (2016).

- Kumar, B. V. et al. Human tissue-resident memory T cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell Rep.* **20**, 2921–2934 (2017).
- Im, S. J. et al. Defining $CD8^+$ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).
- Miller, B. C. et al. Subsets of exhausted $CD8^+$ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
- Balin, S. J. et al. Human antimicrobial cytotoxic T lymphocytes, defined by NK receptors and antimicrobial proteins, kill intracellular bacteria. *Sci. Immunol.* **3**, eaat7668 (2018).
- Thommen, D. S. et al. A transcriptionally and functionally distinct $PD1^+$ $CD8^+$ T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat. Med.* **24**, 994–1004 (2018).
- Sun, Q., Hao, Q. & Prasad, K. V. Nuclear long noncoding RNAs: key regulators of gene expression. *Trends Genet.* **34**, 142–157 (2018).
- Delmas, V., Stokes, D. G. & Perry, R. P. A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc. Natl. Acad. Sci. USA* **90**, 2414–2418 (1993).
- Gaide, O. et al. Common clonal origin of central and resident memory T cells following skin immunization. *Nat. Med.* **21**, 647–653 (2015).
- Simoni, Y. et al. Bystander $CD8^+$ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
- Scheper, W. et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat. Med.* **25**, 89–94 (2019).
- Mariathasan, S. et al. TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
- Fehrenbacher, L. et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387**, 1837–1846 (2016).
- McDermott, D. F. et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat. Med.* **24**, 749–757 (2018).
- Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
- Araujo, J. M. et al. Effect of CCL5 expression in the recruitment of immune cells in triple negative breast cancer. *Sci. Rep.* **8**, 4899 (2018).
- Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
- Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486–499 (2015).
- Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* **27**, 450–461 (2015).
- Khan, O. et al. TOX transcriptionally and epigenetically programs $CD8^+$ T cell exhaustion. *Nature* **571**, 211–218 (2019).
- Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).
- Sade-Feldman, M. et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* **175**, 998–1013 (2018).
- Yan, Y. et al. CX3CR1 identifies PD-1 therapy-responsive $CD8^+$ T cells that withstand chemotherapy during cancer chemimmunotherapy. *JCI Insight* **3**, e97828 (2018).
- Schumacher, T. N. & Scheper, W. A liquid biopsy for cancer immunotherapy. *Nat. Med.* **22**, 340–341 (2016).
- Hogan, S. A. et al. Peripheral blood TCR repertoire profiling may facilitate patient stratification for immunotherapy against melanoma. *Cancer Immunol. Res.* **7**, 77–85 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Reagents and antibodies

Antibodies used in FACS sorting were: EpCAM: Biolegend 9C4 antibody, cat. no. 324208, lot B198775, fluorochrome APC. CD45: Biolegend 2D1 antibody, cat. no. 368516, lot B264243, fluorochrome APC/Cy7. CD3: BD Biosciences SK7 antibody, cat. no. 564001, lot 6109971, fluorochrome BUV395.

Software versions

Data were collected using Cell Ranger software (10x Genomics) v.2.2.0³². Data were analysed using Perl v.5.18.2, R v.3.6.0, and the following packages and versions in R for analysis: Seurat, 3.0.2³³; SingleR, 1.0.1³⁴; RankProd, 3.11.0³⁵; GSEABase, 1.47.0; limma, 3.41.15³⁶; annotate, 1.63.0; homologue, 1.4.68.19.3.27; mouse4302.db, 3.2.3; cocor, 1.1-3³⁷; iNEXT, 2.0.19³⁸; and survival, 2.44-1.1³⁹. Two-dimensional gene expression maps were generated using coordinates from the UMAP algorithm^{40,41}, as implemented in the umap-learn module v.0.3.10 in Python v.3.6.0, run under R using reticulate v.1.13. Figures were produced using the following packages and versions in R: colorspace, 1.4-1⁴²; RColorBrewer, 1.1-2; pheatmap, 1.0.12; and superheat, 0.1.0⁴³. External datasets were obtained using GenomicDataCommons, 1.9.0⁴⁴ and GEOquery, 2.53.0⁴⁵.

Human research participants

Patient samples were procured from commercial vendors (Discovery Life Sciences, iSpecimen Inc., Avaden BioSciences and TriMetis Life Sciences) from adult patients undergoing resection surgery. We complied with all ethical standards of the Roche Ethics Committee. Informed consent was obtained from all participants.

Tissue dissociation

Fresh surgical samples were separated into tumour and NAT compartments by the reviewing pathologist and shipped overnight to our institution. Upon arrival, samples were rinsed with PBS until no traces of blood were visually detected. Subsequently, samples were digested with a combination of collagenase D (0.5 mg ml⁻¹) and DNase (0.1 mg ml⁻¹) for 15 min at 37 °C with gentle shaking. Subsequently, samples were subjected to a gentle MACS dissociator (Miltenyi Biotec), followed by an additional 10 min incubation at 37 °C.

Peripheral blood mononuclear cell isolation

Peripheral blood mononuclear cells from patient blood samples were isolated using 50-ml Leucosep tubes (Greiner Bio-One International) and Ficoll-Paque PLUS (GE Healthcare). Whole blood drawn into sodium heparin blood collection tubes were diluted three times with PBS without calcium or magnesium (Lonza). Diluted cell suspensions were carefully layered on Leucosep tubes and centrifuged for 15 min at 800g at room temperature. Interphase-containing peripheral blood mononuclear cells were collected and washed with PBS and subsequently centrifuged for 10 min at 250g at room temperature before further processing.

FACS sorting

Following tissue enzymatic dissociation of tissues, single-cell suspensions were subjected to antibody staining with anti-EPCAM, anti-CD45, and anti-CD3. Cells were purified by fluorescence-activated cell sorting (FACS) on a Becton Dickinson FACS Aria Fusion cell sorter equipped with four lasers (405 nm, 488 nm, 561 nm and 638 nm). A 70-µm nozzle running at 70 psi and 90 kHz was used as the setup for each sort session. FACSDiva (v.8.0.1) and FlowJo (v.10) were used to collect and analyse the flow cytometry data. Before gating on fluorescence, single cells were gated using forward scatter (FSC-A) and side scatter (SSC-A) (for intact cells) and SSC-W/SSC-H and FSC-W/FSC-H (to ensure that only singlets were sorted). FACS gates were drawn to include only live single cells based on Calcein Blue AM⁺ and 7-AAD (Thermo Fisher Scientific).

Further gates were drawn to arrive at CD3⁺CD45⁺EpCAM⁻ (for CD3⁺ selected samples) or CD45⁺EpCAM⁻ cells (for CD45⁺ selected samples). Boundaries between positive and negative cell fractions were drawn based on single-colour stains. An example gating strategy is shown in Supplementary Methods.

Single-cell RNA-seq and TCR V(D)J clonotype profiling

Sample processing for single-cell gene expression (scRNA-seq) and T cell receptor V(D)J clonotypes (scTCR-seq) was done using the Chromium Single Cell 5' Library and Gel Bead Kit (10x Genomics), following the manufacturer's user guide. Cell density and viability of FACS-sorted CD3⁺ T cells or CD45⁺ cells from tumours, NAT and blood were determined by Vi-CELL XR cell counter (Beckman Coulter). All of the processed samples had cell viability above 90%. The cell density was used to impute the volume of single cell suspension needed in the reverse transcription master mix, aiming to achieve approximately 6,000–10,000 cells per sample. After Gel Bead-in-Emulsion reverse transcription (GEM-RT) reaction and clean-up, a total of 14 cycles of PCR amplification were performed to obtain sufficient cDNAs used for both RNA-seq library generation and TCR V(D)J targeted enrichment followed by V(D)J library generation. TCR V(D)J enrichment was done per manufacturer's user guide using Chromium Single Cell V(D)J Enrichment Kit, Human T cell (10x Genomics). Libraries for RNA-seq and V(D)J were prepared following the manufacturer's user guide (10x Genomics), then profiled using Bioanalyzer High Sensitivity DNA kit (Agilent Technologies) and quantified with Qubit (Thermo Fisher Scientific) or Kapa Library Quantification Kit (Kapa Biosystems). Single-cell RNA-seq libraries were sequenced in one lane of HiSeq4000 (Illumina). Single-cell TCR V(D)J libraries were tagged with a sample barcode for multiplexed pooling with other libraries, sequenced in both lanes of a HiSeq2500 machine (Illumina) using Rapid Run mode, and then demultiplexed. All sequencing was done according to the manufacturer's specification (10x Genomics).

Processing of scTCR-seq data

TCR-seq data for each sample was processed using Cell Ranger software with the command 'cellranger vdj' using a custom reference set of 30,727 genes, based on human reference genome GRCh38 and RefSeq gene models, and provided in our software package (see 'Code availability'). For each sample, Cell Ranger generated an output file, filtered_contig_annotations.csv, containing TCR α-chain and β-chain CDR3 nucleotide sequences for single cells that were identified by barcodes.

Although Cell Ranger software groups cells according to CDR3 nucleotide sequences to obtain clonotypes separately for each sample, our analysis depends on identifying shared clonotypes across samples. We therefore re-grouped clonotypes across the tumour, NAT and blood samples for each patient, requiring that they share all reported α-chain and β-chain CDR3 consensus nucleotide sequences in common. We tested our procedure on each sample individually and confirmed that it grouped clonotypes consistently with the original clonotypes from the Cell Ranger software.

Clonotypes were assigned a tissue expansion pattern based on their clone sizes in NAT and tumour as follows. Clonotypes having one cell in NAT but none in tumour were called NAT singletons, while clonotypes having more than one cell in NAT but none in tumour were called NAT multiplets. Conversely, clonotypes having one cell in tumour but none in NAT were called tumour singletons, and clonotypes having more than one cell in tumour but none in NAT were called tumour multiplets. Clonotypes having at least one cell in NAT and at least one cell in tumour were called dual-expanded clones.

Clonotypes from blood samples were classified according to the above criteria if they had any cells in NAT or tumour. The remaining clonotypes, which had cells only in blood and none in NAT or tumour, were called blood singletons or blood multiplets if they had one cell

Article

in blood or more than one cell, respectively. To study the overlap of clonotypes between blood and tissue, we also classified clonotypes from blood samples as either blood non-expanded or expanded on the basis of their blood clone sizes being one or more than one, respectively, without regard to their NAT or tumour clone sizes.

Processing of scRNA-seq data

scRNA-seq data were processed with Cell Ranger software. Illumina base call (BCL) files were converted to FASTQ files with the command 'cellranger mkfastq'. Expression data were processed with 'cellranger count' on the pre-built human reference set of 30,727 genes. Cell Ranger performed default filtering for quality control, and produced for each sample a barcodes.tsv, genes.tsv, and matrix.mtx file containing counts of transcripts for each gene in each single cell, which we parsed using a script in R to create a count matrix for each sample. Single cells were identified by barcodes that were consistent with those from the scTCR-seq dataset. In total, scRNA-seq data were obtained on 541,863,094 transcripts in 200,626 immune cells.

Cluster analysis of combined immune cells

Count matrices for each sample were processed using Seurat using the SCTransform procedure with default parameters to perform a regularized negative binomial regression based on the 3,000 most variable genes. Normalized datasets for the tumour, NAT and blood (if available) samples for each patient were combined using the FindIntegrationAnchors and IntegrateData functions in Seurat with the default value of 30 dimensions.

The resulting datasets for the 200,626 immune cells from all 14 patients were then combined using the FindIntegrationAnchors and IntegrateData functions in Seurat with the default value of 30 dimensions. This process required 130 GB of computer memory. The integrated dataset was scaled and processed under principal components analysis using the ScaleData and RunPCA functions in Seurat. Two-dimensional map coordinates were generated using the RunUMAP procedure. Cluster analysis was performed using the FindNeighbours procedure and the FindClusters procedure at a resolution of 1.6, larger than the default value of 0.8, in order to obtain a finer resolution of cell subtypes, and yielded 33 clusters of both T and non-T cells. The FindMarkers procedure in Seurat was run on each cluster to obtain biomarker genes upregulated in that cluster. The default value of 0.25 for logfc.threshold was found to generate fewer than 20 genes for some clusters, so the procedure was run with a value of 0.10, at the expense of greater compute time.

Computational separation of T and non-T cells

Clusters identified from all 200,626 immune cells were characterized by their mean expression of *CD3E* and their fraction of cells having a clonotype identified from the scTCR-seq assay (Extended Data Fig. 1a). T cell clusters were identified as having high levels of both measures, and non-T cell clusters were those that had low levels of both measures.

Some clusters had intermediate values and their biological meaning was inferred by their biomarker genes from the FindMarkers procedure (Supplementary Table 2). Cluster 28 (Extended Data Fig. 1a, c) was determined to represent plasma cells based on exclusively high expression of *MZB1* (encoding pERp1, plasma cell-induced resident endoplasmic reticulum protein)^{46,47}. Cluster 17 was determined to represent natural killer cells, which had gene expression similar to effector T cells, with cytotoxic biomarkers, such as *NKG7*, *GZMB* and *PRF1*. However, the origin of cells in this cluster was heavily biased towards CD45-selected samples, and the fraction of TCR clonotypes was much lower than for other T cell clusters. The 20% of cells with TCR clonotypes in this cluster were thought to be from NKT cells having gene expression similar to that of natural killer cells. Cluster 29 was determined to represent macrophages expressing TCR receptors, a phenomenon that has been observed by others^{48,49}. The identity of this

cluster as representing macrophages was evidenced by high expression of *SAT1*, which is increased in granulocytes and macrophages, but not in T, B or natural killer cells, as displayed in the Human Protein Atlas⁵⁰ at <http://proteatlas.org/ENSG00000130066-SAT1/tissue>. Clusters 11, 16, 19, 21 and 27 with intermediate *CD3E* expression but clonotype fractions above 40% were determined to represent T cells (Extended Data Fig. 1a, b).

In support of this separation process, Extended Data Fig. 1i shows that CD3-selected samples had relatively few cells classified as non-T cells, and that CD45-selected samples all contributed to the collection of T cells. Furthermore, subsequent cluster analyses of T and non-T cells separately, as described below, confirmed our separation process, resulting in no outlier clusters that were classified inappropriately.

Cluster analyses of T and non-T cells

On the basis of the above separation of T and non-T cells, the entire process of data transformation, integration and cluster analysis was repeated on the two divisions separately. The procedure was the same, with two differences. First, the parameter features.to.integrate = all.features was given to IntegrateData, where all.features was the intersection of genes from the SCTransform procedure across all patients, in order to maximize the genes stored by Seurat into the integrated assay, rather than limiting them to the genes stored as var.features. Second, the resolution for the FindClusters procedure was set to its default value of 0.8.

One cluster of 3,182 T cells was found to have two discrete components in the UMAP representation, as well as a portion of 60 cells that clustered with a mixture of cells from other clusters (the region to the left 4.6b-T_{reg} in Fig. 2a). The latter cells were considered possible contaminating or low quality cells and assigned no cluster. A division of the remaining cells was made provisionally, applying a *k*-means algorithm on the UMAP two-dimensional coordinates using the *k*-means procedure in R. Because subsequent analyses revealed distinct gene expression patterns for the subclusters, the division was therefore kept, and constituted the clusters reported as '8.4-Chrom' and '8.6-KLRB1'.

Identification of markers

To characterize each cluster, we applied both the FindMarkers procedure in Seurat, which identified markers using log fold changes of mean expression, as well as a logistic regression procedure, which identified biomarkers that separated cells of a given cluster from other cells. The logistic regression procedure was performed on each gene within each patient for each cluster using the SCTransform assay data, and yielded a signed deviance value, where the sign depended on the coefficient for the gene expression term, as well as a *t*-statistic. For each cluster, signed deviance values were combined across patients using the Rank Product procedure in the RankProd package in R, and using the resulting one-sided *q*-values for ranking genes.

Gene set enrichment analysis

Gene sets from MSigDB^{51,52} v.7.0 were downloaded as Entrez gene identifiers in GMT format from <http://software.broadinstitute.org/gsea/downloads.jsp> and parsed using the getGmt procedure from the GSEA-Base package in R. For each cluster from the cluster analysis of T cells and each of the 22,596 gene sets, we performed a one-sided *z*-test on the *t*-statistics from the logistic regression procedure in 'Identification of markers' section. This statistical test was previously proposed⁵³ as an alternative to enrichment scores, provided that genes are given a value that has zero mean and unit variance, which should be satisfied theoretically by *t*-statistics and verified by inspection of quantile-quantile plots. To avoid spurious results, we considered only gene sets with five or more genes matching those in our scRNA-seq dataset. Inspection of results showed high ranking gene sets from clusters '8.4-Mitosis' and '8.5-Chrom' against Gene Ontology gene sets corresponding to

biological processes, and '3.1-MT' against the gene set for mitochondrial genes, as shown in Extended Data Fig. 4g.

Comparison of clusters with external single-cell gene signatures

We computed reference gene signatures from the scRNA-seq data of the datasets for Guo et al.³, Zhang et al.⁴ and Yost et al.⁵. The first two datasets had count matrices that were already normalized by library sizes in the form of transcripts per million (TPM) matrices, and we computed this same normalization for the third dataset. For each dataset, we considered only genes with at least one non-zero TPM value across all cells, and computed the \log_2 -transformed expression value $\log_2(\text{TPM} + 1)$, where 1 represents a pseudocount. For each cluster, based on the metadata from the original analyses, we obtained robust centroids by performing a 10% trimmed mean on log-transformed expression values for each gene across cells in the cluster. We used these centroids as reference gene signatures for the R package SingleR, which assigned single-cell assignments to each cell in our dataset. Gene expression measurements from our dataset were the cell counts computed for each sample before integration, again considering only genes with at least one non-zero TPM value across all cells in the sample and computing the \log_2 -transformed expression value $\log_2(\text{TPM} + 1)$. We cross-tabulated counts of cells based on both our internal cluster assignments and the external assignments computed by SingleR, and normalized the resulting matrix first by the external assignments and then by our internal assignments to generate heat maps for Extended Data Fig. 4a.

Display of gene expression

Bar plots in Extended Data Fig. 4b are computed using the scaled data from the integrated assay using Seurat. Because the integrated assay requires that genes be selected as variable in each of the component datasets, some genes could not be depicted this way. In particular, the *PDI* gene was not included in the integrated assay, so its expression in Extended Data Fig. 4d is taken from counts from the original RNA data.

Analysis of data from Guo et al.³

Single-cell RNA-seq counts were obtained using the GEOquery library in R and running the command `getGEOsupFiles` for GSE99254. Expression data were obtained over 23,459 genes in 12,348 cells. Metadata consisting of the patient, cluster assignment, and tissue source for each cell were downloaded from the Gene Expression Omnibus (GEO) project GSE99254 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99254>). TCR sequences and clonotype groupings were obtained from supplementary table 2 of Guo et al.³. TCR sequences were obtained for 10,202 cells in 7,398 distinct clonotypes.

Analysis of data from Zhang et al.⁴

Single-cell RNA-seq counts were obtained using the GEOquery library in R and running the command `getGEOsupFiles` for GSE108989. Expression data were obtained over 23,459 genes in 11,140 cells. Metadata consisting of the patient, cluster assignment, and tissue source for each cell were downloaded from GEO project GSE108989 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989>). TCR sequences and clonotype groupings were obtained from supplementary table 4 of Zhang et al.⁴. TCR sequences were obtained for 9,878 cells in 7,274 distinct clonotypes.

Signature for resident memory T cells

Genes were identified as being positively associated (13 genes) or negatively associated (18 genes) with resident memory T cells, based on Fig. 3a from a published study of tissue-resident versus circulating T cells⁸. The *D4S234E* gene was identified as a synonym for *NSG1* (Entrez ID 27065). Genes were matched against those in the integrated assay from Seurat, showing an overlap with four positively associated genes (*CRTAM*, *ITGAE*, *DUSP6* and *RGS1*) and three negatively associated ones

(*FAM65B*, *STK38* and *KLF2*). Although only a few genes showed overlap, the published study indicates that genes in this signature share highly correlated gene expression. The expression of each gene was converted to z-scores by subtracting the mean expression across samples and dividing by their standard deviation, and we computed a positive signature score for each T cell as the mean z-score from the four positively associated genes and a negative signature score similarly from the three negatively associated genes. The overall score was the difference between the positive and negative signature scores.

Signature for terminal versus stem-like exhaustion

Microarray data, in the form of \log_2 -transformed robust multichip average (RMA) normalized intensities, were obtained from samples GSM2227309 to GSM2227314 from GEO project GSE84105⁹ using the `getGEO` procedure of the GEOquery package in R. Differentially expressed probes between the three CXCR5⁺ mouse samples and the three TIM3⁺ mouse samples were found using the `lmFit`, `contrasts.fit`, `eBayes`, and `topTable` procedures in the `limma` package in R, at a threshold \log_2 -transformed fold change of three or more. These probes were assigned to genes using the `annotate` and `mouse4302.db` packages, and then translated into human equivalents using the `homologene` package in R. The resulting 58 CXCR5-associated genes and 42 TIM3-associated genes were compared against our scRNA-seq dataset using the scaled data array from the integrated assay computed by Seurat, yielding 12 CXCR5-associated genes (*GPR183*, *CCR6*, *CRTAM*, *SLFN5*, *TNFRSF25*, *JUN*, *IFI16*, *SLAMF6*, *XCL1*, *IL7R*, *EMB* and *SATB1*) and 9 TIM3-associated genes (*ADAM8*, *PRDMI*, *CDKN2A*, *CCL4*, *CD7*, *AHR*, *GPR56*, *GZMA* and *CISH*) in common. The expression of each gene was converted to z-scores by subtracting the mean expression across samples and dividing by their standard deviation, and we computed for each T cell a CXCR5 signature score as the mean z-score from the CXCR5-associated genes and a TIM3 signature score similarly from the TIM3-associated genes. The overall score for terminal exhaustion versus stem-like exhaustion was the difference between the TIM3 and CXCR5 signature scores.

Integration of scRNA-seq and scTCR-seq data

One issue in integrating scRNA-seq and scTCR-seq data is that the clonal expansion pattern information from TCR-seq is applied to clones, whereas cluster assignments from RNA-seq are applied to individual cells, based on their transcriptional profiles. Therefore, a clone may contain a diverse set of cells with different clusters (Extended Data Fig. 4h, i). To integrate the data, we assigned tissue and blood expansion patterns from clones to their constituent cells. Clones were assigned a primary cluster, based on the cluster with the largest representation of cells in the clone. In cases of ties, in which the two largest representative clusters had equal counts, we assigned no primary cluster to that clone.

Another consideration is that it was possible for cells to be assigned to the non-T cell category based on the scRNA-seq assay, but given a clonotype based on the scTCR-seq assay. For analyses of scTCR-seq data alone, such cells would be included on the basis of their clonotype. For analyses of scRNA-seq data alone, such cells would be excluded for lacking a T cell cluster. However, for integrated analyses of scRNA-seq and scTCR-seq data, the answer is unclear. As indicated in Extended Data Fig. 1a, the assignment of clonotypes to non-T cells was relatively uncommon in our dataset. In this study, we included such cells in integrated analyses with a special cluster of 'non-T', computed primary clusters as described above, and then assigned no primary T cell cluster to clones that had ties for the two largest representative clusters (with one potentially being non-T) or had non-T as the primary cluster. Cells were then excluded from the integrated analysis if their parent clone lacked a primary T cell cluster. The result of this procedure was to include non-T cells with a clonotype at the outset, but then exclude most of them (including all singletons) because their primary cluster was non-T cells.

Analysis of data from Yost et al.⁵

scRNA-seq counts and single-cell TCR sequences were obtained using the GEOquery library in R and running the command `getGEO` Supp-Files for GSE123813. Metadata consisting of the patient, treatment status, and cluster assignment for each cell were obtained from the corresponding authors for Yost et al.⁵. We analysed expression data that had matching metadata, resulting in 33,106 cells measured on 23,309 genes. Single-cell TCR sequences were available on 28,371 cells with matching metadata. Bulk TCR-seq sequences and template counts were obtained from the ImmuneACCESS Web portal by Adaptive Biotechnologies (<https://clients.adaptivebiotech.com/pub/yost-2019-natmed>).

The scTCR-seq data were provided as TCR- α and TCR- β nucleotide and amino acid sequences without assigned clonotypes. We therefore created clonotype groupings from pre- and post-treatment tumour samples by matching all available TCR chains between cells of a given patient, using a procedure analogous to our re-grouping of tumour, NAT and blood clonotypes for our own dataset. Among the 11 patients in the dataset, this procedure identified 17,369 distinct clonotypes. Patient bcc.su003 had no clonotype matches between pre-treatment and post-treatment tumour samples, and could not be analysed further. The remaining patients had pre-treatment and post-treatment bulk TCR-seq data from blood for five and six patients, respectively, with patient bcc.su002 having no pre-treatment blood sample.

Bulk TCR-seq clonotypes were linked to tumour clonotypes by matching the single TCR β -chain to any of the TCR β -chains from a clonotype in the scTCR-seq data (Extended Data Fig. 7a). The bulk TCR-seq data reports an 87-bp rearrangement sequence from each TCR β -chain that includes the CDR3 sequence, so the scRNA-seq CDR3 (sc-CDR3-nt) was required to match exactly at the nucleotide level to the corresponding substring (bulk-CDR3-nt) in the bulk TCR-seq rearrangement. This alignment process was facilitated by the fact that the immunoSEQ output includes the amino acid CDR3 (bulk-CDR3-aa) sequence, which we used to identify the corresponding bulk-CDR3-nt. We could then create a hash table of all sc-CDR3-nt sequences and use that for lookup from a given bulk-CDR3-nt sequence.

Because scTCR-seq clonotypes can contain TCR α -chains as well as multiple TCR β -chains, the potential exists for a many-to-one correspondence between a scTCR-seq clonotype from tumour and a bulk TCR-seq clonotype from blood. For analyses requiring clones to be classified by the presence or absence of a bulk TCR-seq match (that is, blood association versus blood independence) (Fig. 3c), all matching clonotypes were considered to have an association with blood. For analyses requiring matching clone sizes between tumour and blood (Fig. 3a, Extended Data Fig. 7b, c), multiple tumour clonotype matches for a given blood clonotype were combined by summing the scTCR-seq clone sizes and resolving multiple primary cell clusters by a majority vote or leaving the primary cell cluster unassigned in case of a tie. For the analysis comparing pre- and post-treatment bulk TCR sequences (Extended Data Fig. 7d), clonotypes were matched when the entire 87-bp rearrangement sequences were identical.

Because bulk TCR-seq measures transcripts rather than cells, we could not discriminate blood-expanded from non-expanded clones, although we could distinguish between blood-independent and blood-associated clones by the absence or presence of a matching transcript. We assumed that if the number of transcripts per cell is relatively uniform, the number of transcripts was approximately proportional to the number of cells. Therefore, instead of measuring peripheral clonal expansion, we measured blood clonal diversity using a Shannon entropy measure, the sum of $p \times \log_2(p)$, over all transcripts with a count of 50 or more. This threshold was implemented because of the heavily skewed distribution of transcript counts, with large numbers of singleton transcripts that potentially introduced noise into the calculations.

Completeness curves for Yost et al.⁵

We used the transcripts field from each immunoSEQ analysis file for bulk TCR-seq from a pre- or post-treatment blood sample as input to the iNEXT procedure in the iNEXT package in R, with 'abundance' as the datatype parameter. The resulting interpolated, observed, and extrapolated values were then plotted using the `m` and `SC` variables in the `iNextEst` object for each sample (Extended Data Fig. 7e).

Comparison of TCR sequences with VDJdb

We obtained the June 2018 (2018-06-04) release of VDJdb⁵⁴ (<https://github.com/antigenomics/vdjb-db/releases>). We analysed the file `vdjb.slim.txt` to extract entries where the TCR species was 'Homo-Sapiens' and the epitope species was 'CMV', 'EBV' or 'InfluenzaA'. We compared each clonotype from our data against the `cdr3` field in the VDJdb database, requiring an exact match from any of the CDR3 amino acid sequences for the clonotype.

Comparison of TCR sequences with TCGA

We obtained a set of 716,720 TCR sequences assembled from 9,142 bulk RNA-seq samples in TCGA using the TRUST algorithm⁵⁵. We obtained those sequences from the authors after gaining authorized access to sequence-level data in TCGA. Each sequence was labelled with a putative CDR3 amino acid sequence in the header of the FASTA file. We compared each clonotype from our data, requiring an exact match from any of the CDR3 amino acid sequences for the clonotype against any substring of the TCGA-derived CDR3 sequence. Each FASTA file was annotated with an aliquot ID, which we linked to a cancer type using the R library `GenomicDataCommons`. We considered a TCGA-derived CDR3 sequence to be virally reactive when it was observed in three or more individuals, all with different cancer types, and no two individuals had the same cancer type.

Gene signatures for tissue expansion patterns

We applied a logistic regression procedure similar to that used for finding characteristic biomarkers of T cell clusters, except that we classified cells in tumour tissue on the basis of their parent clones being tumour singletons, tumour multiplets, or dual-expanded clones. The logistic regression procedure was performed on each gene within each patient for each tissue expansion pattern using the scRNA-seq assay data to test all available genes. The library size was included as a covariate to help remove artefacts related to sequencing depth. Each logistic regression fit yielded a t -statistic for a given gene in a given tissue expansion pattern. For each tissue expansion pattern, t -statistics were combined across patients using the Rank Product procedure in the RankProd package in R, and using the resulting one-sided q -values for ranking genes.

A total of 30 genes were selected for each gene expression signature. Because gene signatures are intended for use in bulk RNA-seq data from tumours, which contain non-T immune cells and non-immune cells, we considered the problem that our genes were selected within a restricted universe of T cells and that expression of our genes by non-T cells could introduce noise. We therefore examined the co-expression patterns of our gene signatures in bulk RNA-seq data from pre-treatment tumours, obtained from three clinical trials, as discussed below (Extended Data Fig. 9a). We computed the Pearson's correlation coefficient of z -scores for each pair of genes in each signature across patients in all three trials, and performed hierarchical clustering of the correlation matrix. We cut the hierarchical clustering at the first level and used the larger of the two groups for the gene signatures. To compute a composite score, we used the t -statistics from the logistic regression signatures as weights. However, the relatively high degree of correlation of gene expression meant that weighting had relatively minor effects, and we obtained similar results by giving equal weighting to genes in our gene signatures.

Analysis of clinical trial data

Bulk tumour RNA-seq data were analysed from three clinical trials: (1) IMvigor210 (National Clinical Trials identifiers NCT02951767 and NCT02108652)¹⁸, a single-arm trial studying locally advanced or metastatic urothelial carcinoma and having 354 patients with bulk RNA-seq data; (2) POPLAR (NCT01903993)¹⁹, a dual-arm trial comparing 93 patients with bulk RNA-seq data from non-small-cell lung carcinoma treated with atezolizumab and 100 patients treated with the chemotherapeutic agent docetaxel; and (3) IMmotion150 (NCT01984242)²⁰, a triple-arm trial of patients with renal cell carcinoma, treated with atezolizumab (86 patients), atezolizumab plus the VEGF inhibitor bevacizumab (88 patients), or the tyrosine kinase inhibitor sunitinib (89 patients). RNA-seq counts were normalized by a trimmed mean of M-values, log₂-transformed, and converted to z-scores for each gene by subtracting the means and dividing by the standard deviation across all patients in the trial. Six patients in IMvigor210 had missing values for PFS, and could not be included in survival analysis. Subsequent clinical analysis was performed using the coxph function from the survival package in R.

PFS hazard ratios were computed for each gene by dichotomizing the expression of that gene based on the value being higher or lower than the median, and then fitting a Cox proportional-hazards model to the censored PFS survival times using the dichotomized variable. To evaluate the overall set of hazard ratios for each gene signature, we applied the gene set test proposed previously⁵³. Specifically, we first confirmed from quantile plots that log-transformed hazard ratios were normally distributed. We then found an intersection of genes from each signature in a given clinical trial by matching the ENTREZ gene IDs. The overall hazard ratio was computed as their geometric mean, equivalent to an arithmetic mean of the log-transformed hazard ratios. A z-statistic was computed as the mean of the log-transformed hazard ratios multiplied by the square root of the number of intersecting genes, and the P value was obtained by referring this value to the normal distribution.

To evaluate combined signature scores, we found an intersection of genes from each signature in a given clinical trial as before. The resulting common genes in each signature were used to calculate an expression score for each patient in each trial as a weighted mean z-score, using mean t-statistics from the logistic regression models to weight each gene. The signature score was dichotomized in each clinical trial based on its value being higher or lower than the median. A Cox proportional-hazards model of this dichotomized variable was fitted to the censored PFS survival data in each arm of each clinical trial, and reported as the hazard ratio and P value of the Wald statistic for each Kaplan–Meier survival plot.

To compute interactions between each signature score and CD8A expression, each score was dichotomized in each clinical trial based on its value being higher or lower than the median. Patients were then classified into one of four groups, depending on the combination of these dichotomized variables. Survival analysis was performed on these groups, with the hazard ratio and P values computed using the dual-low group as the control. Interactions between the tumour multiplet and dual-expanded signatures were analysed analogously using dichotomized values for these signature scores.

Statistical analyses

Aside from the clinical analyses described above, the following statistical tests were used in this study. (1) Chi-square tests were performed using the chisq.test function in R. (2) Post hoc analyses of cells in a contingency table for over-representation were performed as previously described⁵⁶ in which a 2 × 2 table is created from the counts in a given cell and the sum of remaining cells in the same row and column. For each cell, a Fisher exact test was applied using the fisher.test function in R with alt = “greater” to test the null hypothesis that the observed/

expected ratio ≤ 1, in which expected counts derive from the product of the row and column marginals. (3) Pearson’s correlation coefficients and associated two-sided P values were computed using the cor.test function in R to test the null hypothesis that the correlation coefficient is zero. (4) Tests of linear regression t-tests were performed using the lm function in R, followed by the summary function to obtain the two-sided P value of the null hypothesis that the regression coefficient is zero. In some cases, transformations of the variables were performed in addition to linear models, to test the null hypothesis that no relationship exists between the variables. (5) t-tests were performed on log-transformed clone sizes using the t.test function in R to obtain the two-sided P value of the null hypothesis that the two groups have equal means. (6) One-sample z-tests were performed using the pnorm function in R to test the null hypothesis that values in a group have mean ≤ 0 (for the GSEA analysis in Extended Data Fig. 4g, with lower tail set to FALSE) or mean ≥ 0 (for the hazard ratio analysis in Fig. 3a). The application of this test for GSEA was previously described⁵³. The value provided to pnorm was the mean of the values times the square root of the number of cases. For the GSEA analysis, in which biological insight was the main goal, no consideration was made for dependency among genes. For the hazard ratio analysis, we computed the correlation matrix R across the 30 genes based on the expression values across the patients in each clinical trial. The effective number of cases was given by the square root of 1^TR1, where ‘1’ indicates a vector of ones. This procedure reduced the effective value of n by 20–25%. (7) Comparisons of correlation coefficients was performed using a Fisher’s z-test using the corcor function from the corcor package in R to test the null hypothesis that two correlation coefficients from separate samples are equal.

No statistical methods were used to predetermine sample size. The experiments were not randomized unless otherwise stated. All figures of UMAP plots randomize the order in which cells are plotted. Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

FASTQ files containing raw reads from the scRNA-seq and scTCR-seq analyses have been deposited with the European Genome-phenome Archive (EGA) under studies EGAS00001003993 and EGAS00001003994, and datasets EGAD00001005464 and EGAD00001005465. These files are available under controlled access upon request to the Data Access Committee, with contact information provided at EGA (<https://www.ebi.ac.uk/ega/home>). Processed output files from Cell Ranger, integrated assay results from Seurat, and metadata with UMAP coordinates, cluster assignments, and clonotypes are available from the NCBI GEO under accession GSE139555.

Code availability

Computer code used to generate the analyses and figures in this paper are provided as a Supplementary File to the NCBI Gene Expression Omnibus (GEO) accession GSE139555.

32. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
33. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
34. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
35. Del Carratore, F. et al. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. *Bioinformatics* **33**, 2774–2775 (2017).
36. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

37. Diedenhofen, B. & Musch, J. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* **10**, e0121945 (2015).
38. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
39. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).
40. McInnes, L. & Healy, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arXiv.org/abs/1802.03426> (2018).
41. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).
42. Zeileis, A. et al. colorspace: a toolbox for manipulating and assessing colors and palettes. Preprint at <https://arXiv.org/abs/1903.06490> (2019).
43. Barter, R. L. & Yu, B. Superheat: an R package for creating beautiful and extendable heatmaps for visualizing complex data. *J. Comput. Graph. Stat.* **27**, 910–922 (2018).
44. Morgan, M. T. & Davis, S. R. GenomicDataCommons: a bioconductor interface to the NCI Genomic Data Commons. Preprint at <https://www.biorxiv.org/content/10.1101/117200v1> (2017).
45. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).
46. Shimizu, Y., Meunier, L. & Hendershot, L. M. pERp1 is significantly up-regulated during plasma cell differentiation and contributes to the oxidative folding of immunoglobulin. *Proc. Natl Acad. Sci. USA* **106**, 17013–17018 (2009).
47. Andreani, V. et al. Cochaperone Mzb1 is a key effector of Blimp1 in plasma cell differentiation and β 1-integrin function. *Proc. Natl Acad. Sci. USA* **115**, E9630–E9639 (2018).
48. Beham, A. W. et al. A TNF-regulated recombinatorial macrophage immune receptor implicated in granuloma formation in tuberculosis. *PLoS Pathog.* **7**, e1002375 (2011).
49. Fuchs, T. et al. Expression of combinatorial immunoglobulins in macrophages in the tumor microenvironment. *PLoS One* **13**, e0204108 (2018).
50. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, 820 (2017).
51. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
52. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
53. Irizarry, R. A., Wang, C., Zhou, Y. & Speed, T. P. Gene set enrichment analysis made simple. *Stat. Methods Med. Res.* **18**, 565–575 (2009).
54. Shugay, M. et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* **46**, D419–D427 (2018).
55. Li, B. et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat. Genet.* **48**, 725–732 (2016).
56. Shan, G. & Gerstenberger, S. Fisher's exact approach for post hoc analysis of a chi-squared test. *PLoS One* **12**, e0188709 (2017).

Acknowledgements We thank the Genentech FACS Core for supporting the prompt sorting of T cells, and S. Kummerfeld for advice and assistance with initial data analyses. We also thank A. Lun for discussions regarding reference gene signatures and labelling of single cells.

Author contributions J.L.G. and P.E.dA. conceived and designed the study. T.D.W. conceived and performed bioinformatic data analysis and generated figures. P.E.dA., D.E.dA.N., E.Y.C., X.D., H.-M.L., A.S.C., K.L.B., H.I., C.P., A.A.-Y., C.T., S.K. and S. Madireddi processed and prepared samples for sequencing and provided valuable discussion and analysis. Y.-J.J.C., L.D.G., S. Madireddi, P.E.dA. and Z.M. optimized and performed single-cell RNA and TCR sequencing and processing. W.E.O'G., S. Mariathasan, M.B., P.C., M.D.T., M.A.H. and I.E. managed biomarker data collection from clinical trials. R. Banichereau conceived and analysed association with clinical data. T.D.W., L.D., R. Bourgon, I.M. and J.L.G. contributed to discussions and writing of the manuscript. All authors reviewed and approved the final version.

Competing interests All authors are employees of Genentech, which develops and markets drugs for profit.

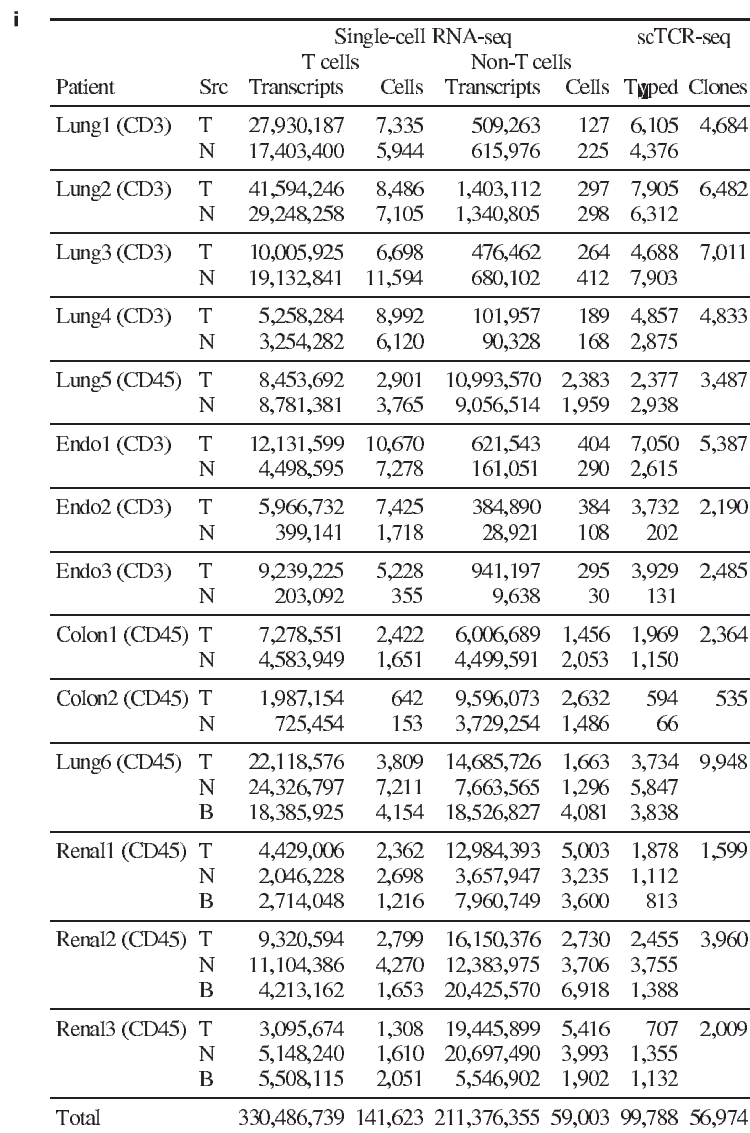
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2056-8>.

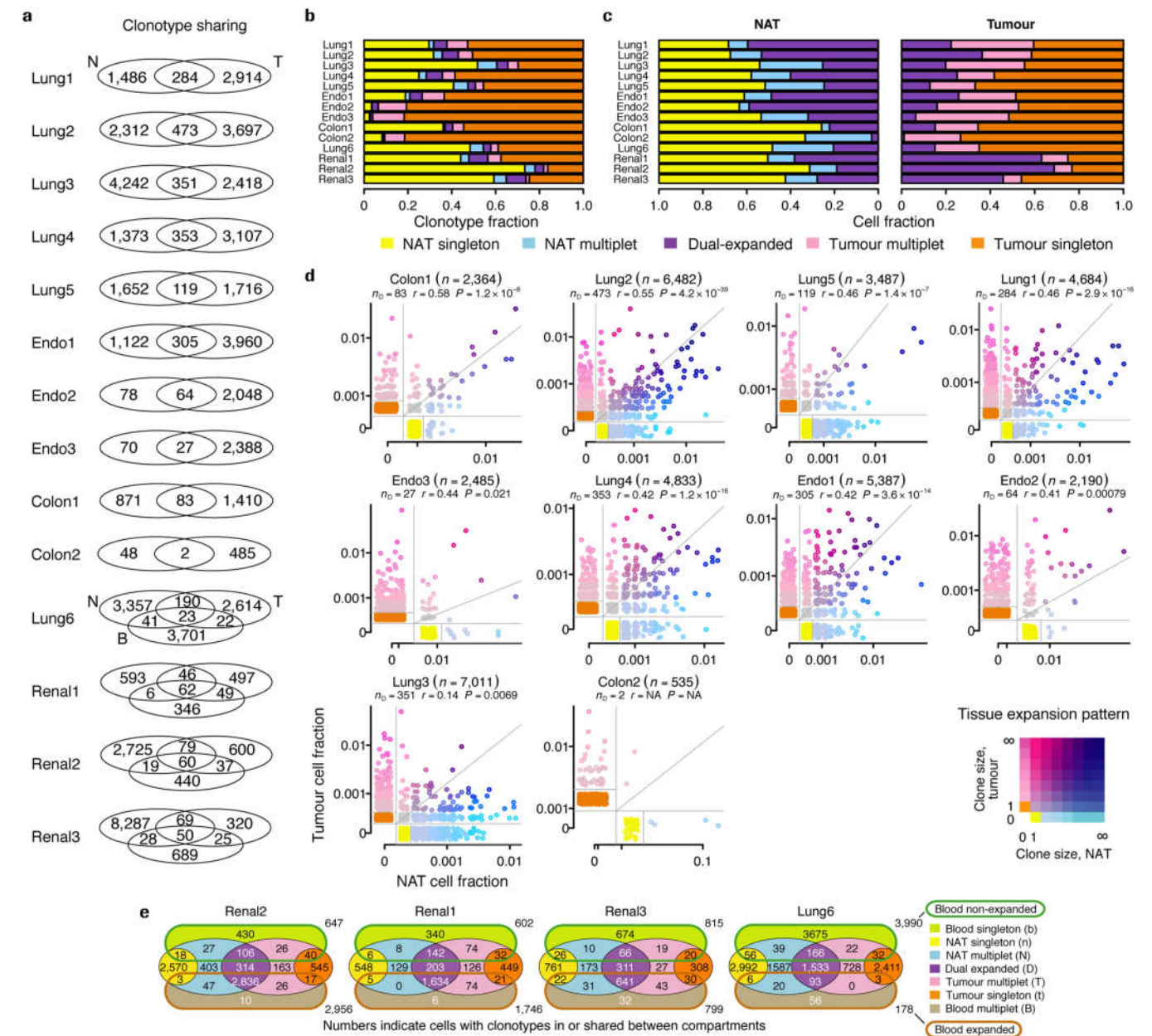
Correspondence and requests for materials should be addressed to T.D.W. or J.L.G.

Peer review information *Nature* thanks Xiang Chen, Xiaole Shirley Liu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

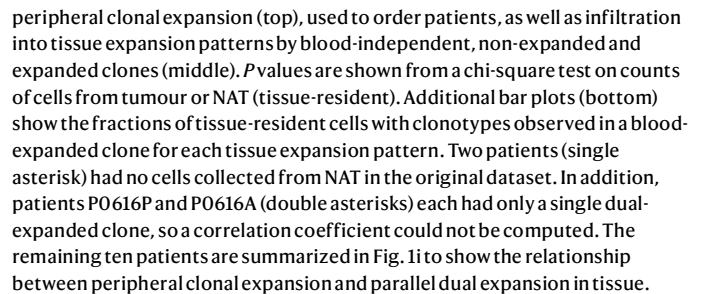


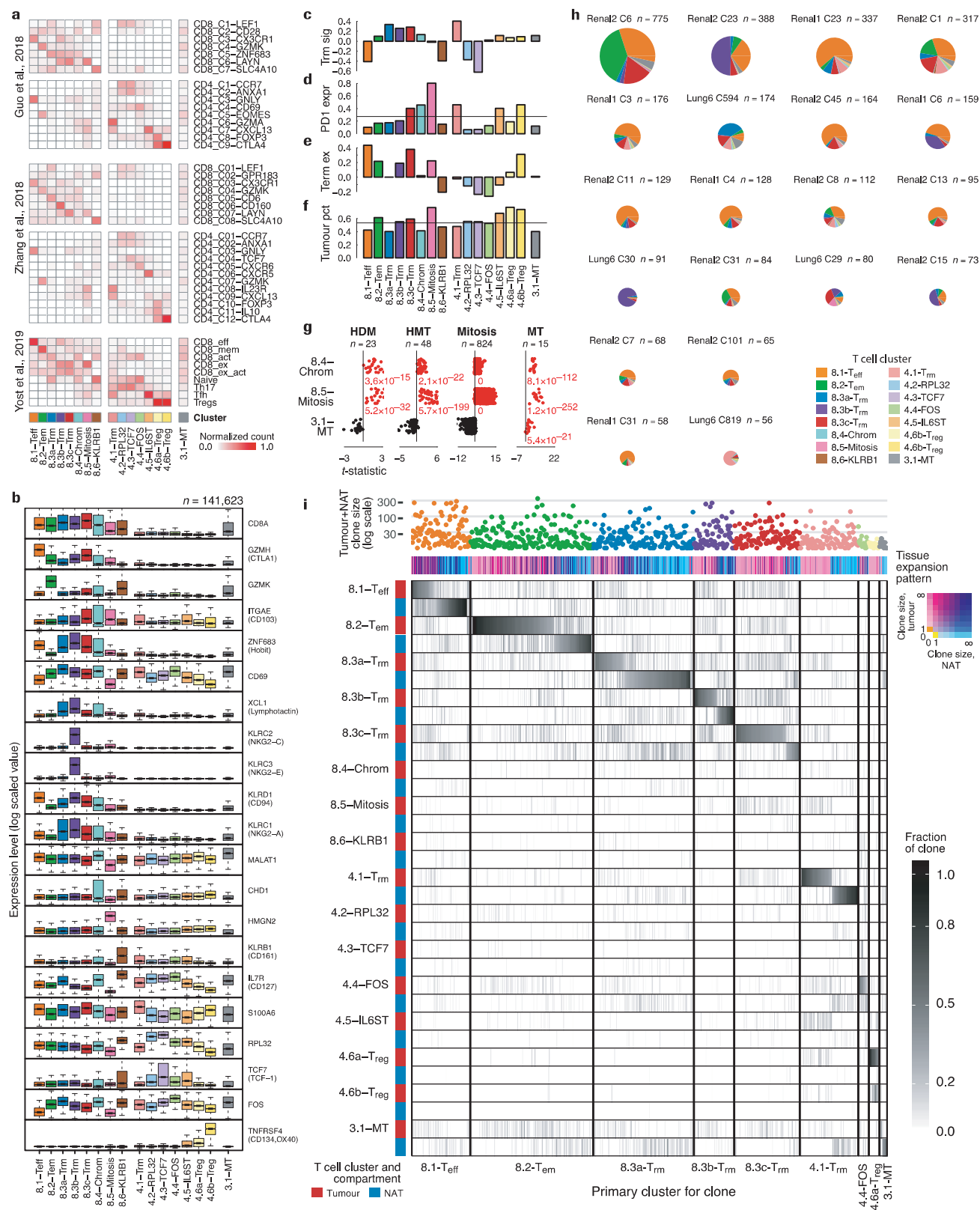
row labelled 'new' match the schema in **h, f**. Mixing of T cells across patients. T cells are mapped by the UMAP algorithm in a subsequent analysis of the T cell division, and are coloured by the patient of origin. Patients are observed to be well mixed across the map, indicating adequate integration of the individual samples. **g**, Clonotype fractions for T cell clusters. Bar plot shows the fraction of cells with a TCR clonotype for each T cell cluster, coloured according to the schema in **d, h**, UMAP plot of non-T cells. Non-T cells from CD45-selected samples are mapped by the UMAP algorithm in a subsequent analysis of the non-T cell division, and are coloured by their cluster assigned by cluster analysis, using the schema in **e**. **i**, Sample statistics. Statistics are provided for the cells in our dataset after separation into T cell and non-T cell categories. Patients are labelled by their cancer type: non-small-cell lung adenocarcinoma (lung), endometrial adenocarcinoma (endo), colorectal adenocarcinoma (colon), and renal clear cell carcinoma (renal). Each patient is annotated by whether the cells were selected by CD3 or CD45, and statistics are given separately for tumour (T), normal adjacent tissue (N), and peripheral blood (B) samples. Numbers indicate the total counts of transcripts and cells from scRNA-seq, as well as the count of cells with clonotypes from scTCR-seq in the column labelled 'typed'. Cells were grouped into distinct clonotypes, and the count of distinct clonotypes is shown for each patient in the column labelled 'clones'.



Extended Data Fig. 2 | Clonotype sharing and tissue expansion patterns. a, TCR sharing across compartments. Venn diagram for each patient shows sharing of TCR clonotypes across compartments. Values indicate the numbers of distinct clonotypes unique to each compartment or shared among compartments in the overlapping oval regions. **b,** Distribution of clones by tissue expansion patterns. Bar plot for each patient shows the fraction of clones having each tissue expansion pattern. **c,** Distribution of cells by tissue expansion pattern. Bar plots for each patient show the distribution of cells in the NAT (left) and tumour (right) compartments according to the tissue expansion pattern of their parent clone. **d,** Clonal expansion in tissue. Scatter plot for each patient shows each distinct clonotype as a dot, with coordinates indicating normalized clone size, or cell fraction, in the NAT and tumour compartments. Dots are coloured by a two-dimensional palette in the bottom right, in which blue shades intensify with increasing NAT clone size, pink shades intensify with increasing tumour clone size, and purple shades intensify

with increasing clone sizes in both compartments. NAT and tumour singletons are indicated by yellow and orange, respectively. Vertical and horizontal grey lines in each scatter plot indicate divisions between absence (clone size of 0 cells) and presence (clone size of 1 or more cells). Diagonal grey lines indicate equal cell fractions in the two compartments. Numerical values in each title indicate the count of distinct clonotypes in each patient. Two-sided P values are from a Pearson's correlation coefficient r on log-transformed clone sizes from the dual-expanded clones (n_D). NA indicates that statistics could not be computed for two or fewer clones. Patients are ordered by decreasing values of r . **e,** TCR sharing by T cells across compartments. Each patient in **a** is represented by a Venn diagram, following the schema in Fig. 1b. Numbers within the Venn diagram regions represent counts of T cells by the tissue and blood expansion patterns of their parent clone. Numbers to the right of each diagram indicate the total number of cells from blood non-expanded and expanded clones, used for computing peripheral clonal expansion in Fig. 1c.



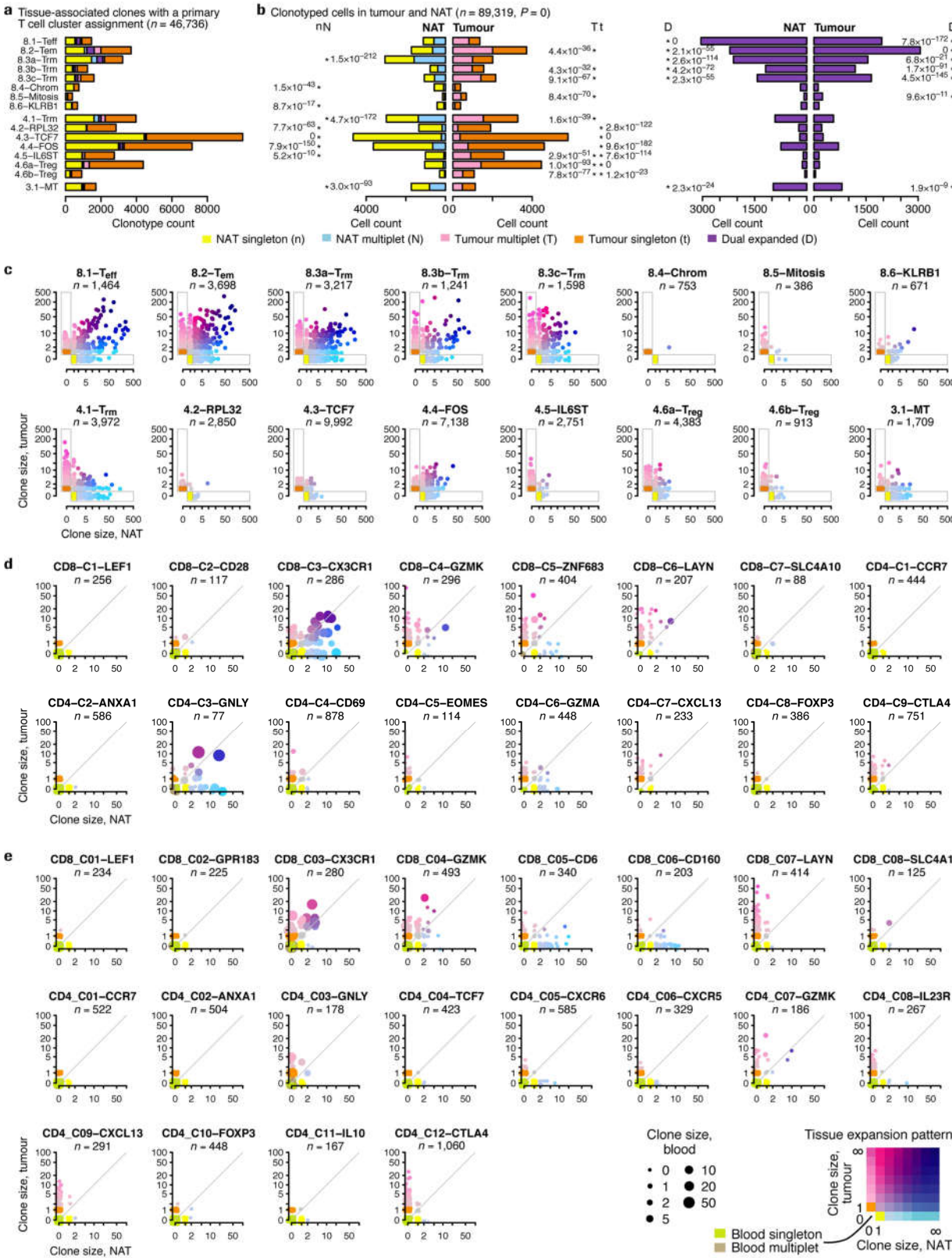


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | T cell subsets and clonal composition. a,

Characterizing clusters of T cell clusters with reference gene signatures. Heat maps show cross-labelling of T cell clusters (columns) to reference gene signatures (rows), taken from the analyses in Guo et al.³, Zhang et al.⁴ and Yost et al.⁵, with intensities indicating normalized frequency. CD8 and CD4 clusters from Guo et al.³ and Zhang et al.⁴ are separated by an extra space to aid visualization. **b**, Expression of selected genes. Box plots show distributions of gene expression on all T cells in the dataset, with cells grouped by their clusters, coloured as in Fig. 2a. Tops and bottoms of boxes indicate interquartile ranges, and lines within boxes indicate medians. Whiskers extend an additional $1.5 \times$ the interquartile range from the median. **c–f**, Characterization of T cell clusters. Bar plots show mean values across clusters of various measures on T cells. 'PD1 expr' denotes expression of *PDI* (**d**); 'Term ex' denotes a published signature of terminal versus stem-like exhaustion⁹ (**e**); 'Trm sig' denotes a published signature of T_{rm} cells⁸ (**c**); and 'Tumour pct' denotes the fraction of cells sampled from tumour versus NAT (**f**). Horizontal lines in **d** and **f** indicate mean values over all cells. **g**, Gene set enrichment analysis for selected clusters and gene sets. The expression of selected gene sets (columns) is shown for clusters (rows) by plotting each gene in the gene set that was assayed in the integrated dataset as a dot according to its t -statistic from a logistic regression

analysis to identify biomarkers for each cluster. Gene Ontology gene sets shown are: histone demethylase activity (HDM); histone methyltransferase activity (HMT); mitotic cell cycle (mitosis); and mitochondrial chromosome genes (MT). A predominance of dots to the right of the vertical line ($t=0$) indicates overexpression of the gene set relative to the expected zero mean. Statistically significant cases of overexpression are shown in red with the associated genes when a one-sided $P < 0.001$ from a one-sample z -test on the t -statistics. **h**, Transcriptional heterogeneity of T cell clones. Each pie chart represents one of the 20 largest clonotypes in this study, as measured by total clone size across tumour, NAT and blood. Each clone represents a set of cells, indicating its total clone size, used to order clonotypes. The area of each pie is proportional to the clone size. Regions of each pie chart indicate the fractions of cells in the given clone assigned to each cluster. **i**, Composition of clones by T cell cluster and compartment. Heat map shows the unit-normalized cellular composition of 770 clones with a tumour + NAT clone size ≥ 10 (columns) across T cell clusters and tumour or NAT compartment (rows). Clones are integrated from all patients and grouped by their primary cluster. Within each primary cluster, clones are ordered to show a gradation of cell fraction from tumour to NAT. Each clone is further characterized by its clone size (top) and tissue expansion pattern (coloured bars above the heat map).

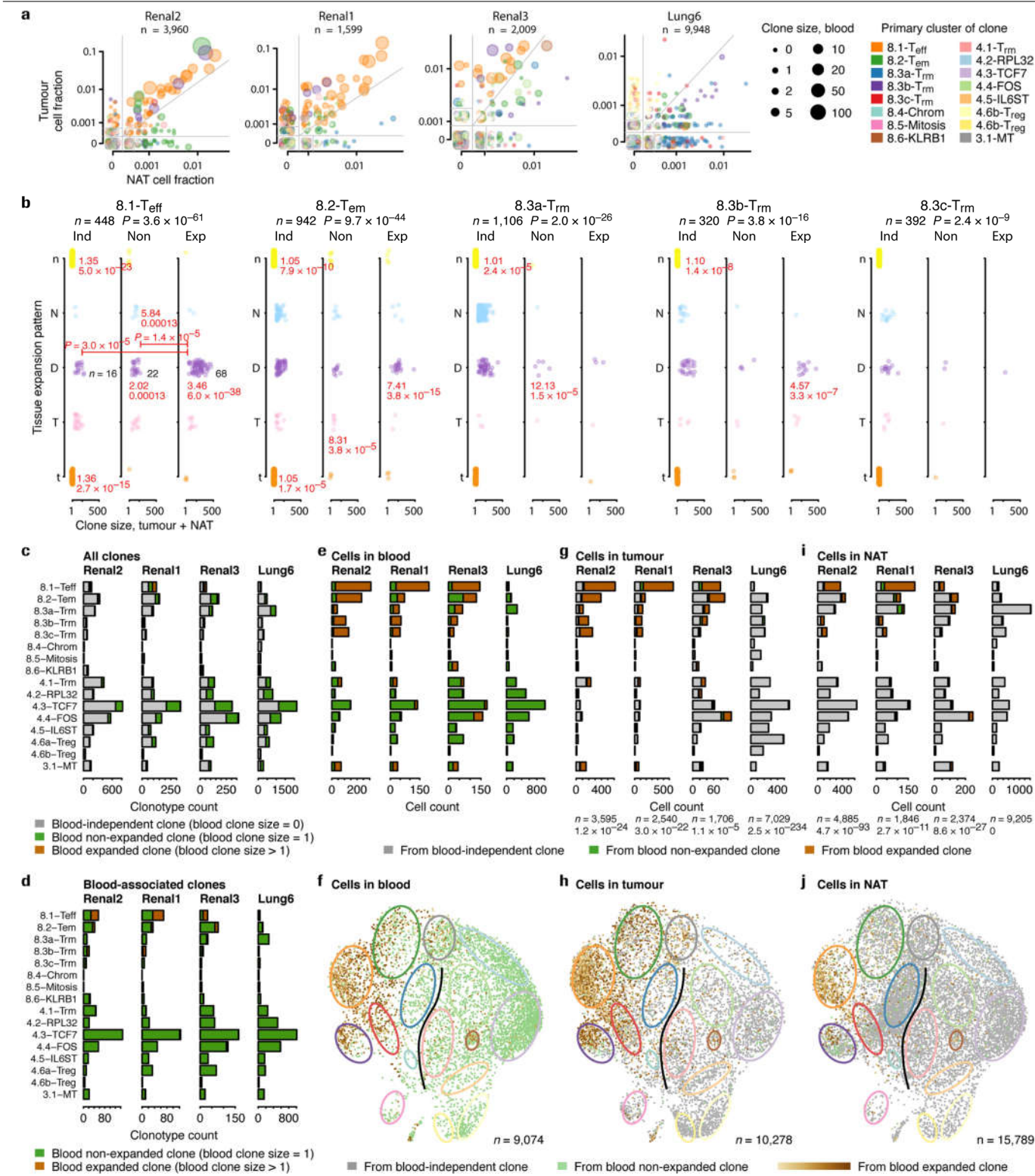


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | T cell subsets and clonal expansion behaviour. a,

Tissue expansion patterns of clonotypes by T cell cluster. Bar plot shows the distribution of tissue-associated clones—having at least one cell in tumour or NAT—and a primary T cell cluster assigned, grouped by primary cluster. Clones in each primary cluster are further divided by their tissue residency pattern. **b,** Tissue expansion patterns of cells by T cell cluster. Bar plots show distributions of T cells in NAT and tumour compartments, grouped by their assigned cluster. The counts in each row, corresponding to a cluster in **a**, comprise all tissue-resident cells—from tumour or NAT—assigned to that cluster. Cell counts are further distinguished by the tissue expansion pattern of their parent clone, with dual expansion shown on the right pair of bar plots, and singletons and multiplets shown on the left. *P* value is from a chi-square test on counts of tissue-resident T cells. Asterisks indicate statistically significant over-representation of the given T cell cluster and tissue expansion pattern, with a one-sided *P* value from a post hoc Fisher exact test on the same counts of tissue-

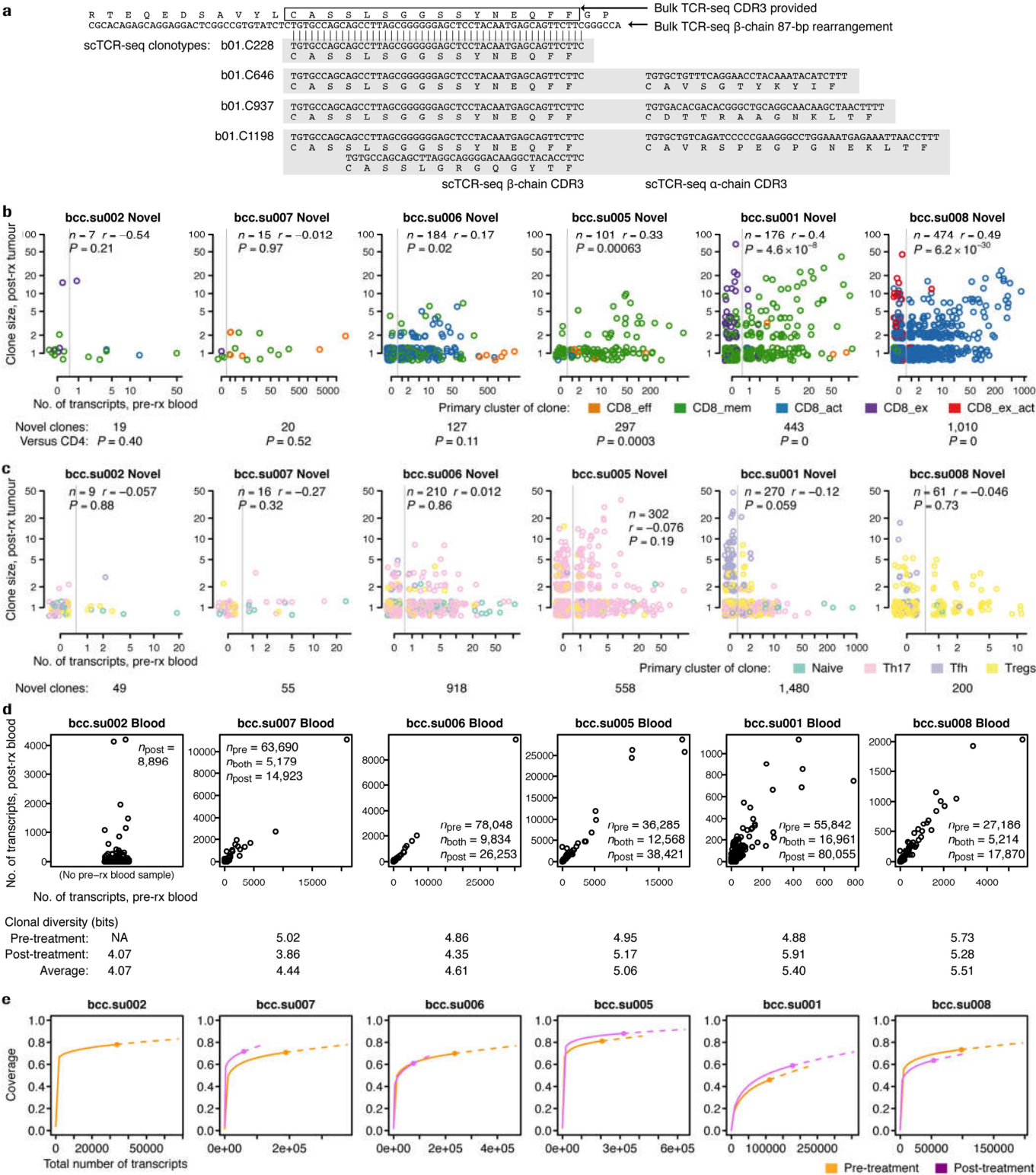
resident T cells as the chi-square test, shown when a Bonferroni-adjusted $P < 0.01$. **c,** Clonal expansion patterns for T cell clusters. Scatter plot for each T cell cluster shows tissue-associated clones with the corresponding primary cluster, integrated from all 14 patients in this study and plotted by their clone sizes in NAT and tumour on logarithmic scales. Dots are coloured by their tissue expansion pattern, as per the two-dimensional palette, except that blood singleton and multiplet clones were not plotted because only four patients had blood samples. **d, e,** Analysis of external datasets. The same methodology of **c** was applied to datasets from Guo et al.³ on 14 patients with non-small cell lung carcinoma (**c**) and Zhang et al.⁴ on 12 patients with colorectal adenocarcinoma (**d**). Clones were grouped according to their primary cluster from the original analyses, and coloured by the two-dimensional palette for tissue expansion pattern at the bottom right of **e**. Blood clone sizes are indicated by dot size, as in **e**.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Tissue and blood expansion patterns and T cell clusters. **a**, Clonal expansion scatter plots by patient. Data from Fig. 1a are shown, except clones are coloured by their primary cluster (see legend). **b**, Clonal expansion by cluster and tissue and blood expansion pattern. The 15 scatter plots in Fig. 2c are represented as vertical sets of strip charts, with each chart showing the clone sizes in tumour plus NAT for clones in each tissue expansion pattern in the scatter plot (abbreviated as n, N, D, T and t, as in Fig. 1b). Strip charts are organized by primary cluster and blood expansion pattern: blood-independent, blood non-expanded and blood-expanded. These one-dimensional representations facilitate the comparison of clone sizes and depiction of statistical results. *P* values are shown from a chi-square test of counts of clones. For each strip chart, the observed/expected ratio and one-sided *P* values are shown in red when a Bonferroni-adjusted $P < 0.01$ from post hoc Fisher exact tests of the same counts of clones as the chi-square test. Additional statistical tests were performed to compare mean clone sizes between the blood-independent, blood non-expanded and blood-expanded categories for each tissue expansion pattern in each cluster. Only two tests had Bonferroni-adjusted $P < 0.01$, shown as bars in the 8.1-T_{eff} dual-expanded category, with two-sided *P* values from a *t*-test on log-transformed clone sizes. **c**, Blood expansion patterns by T cell cluster. Bar plots show the numbers of clones in each of the four patients with a blood sample, with clones grouped by

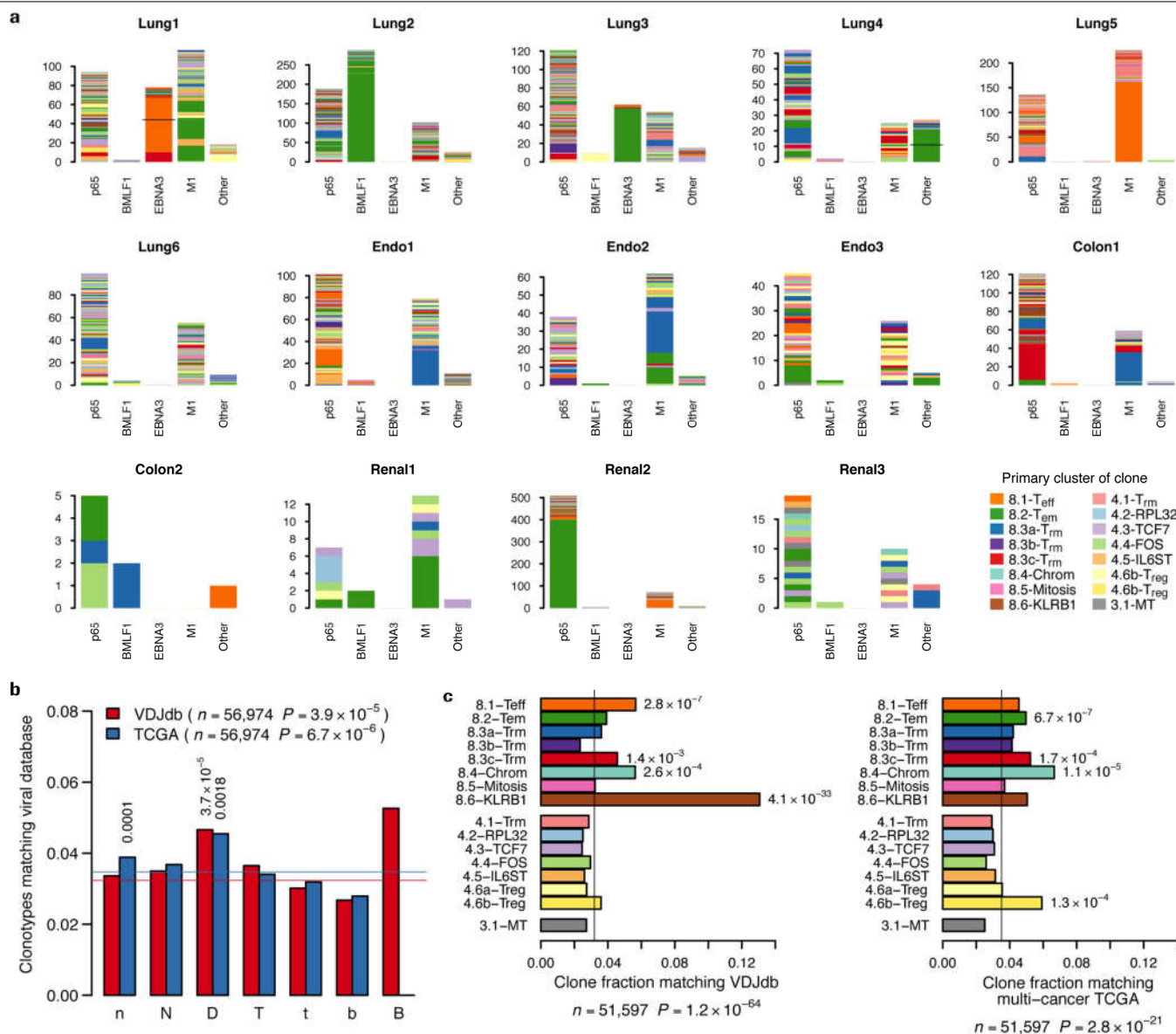
their primary cluster and further divided by their blood clone size as being blood-independent, blood non-expanded or blood-expanded. **d**, Blood-associated expansion by T cell cluster. As in **c**, except blood-independent clones are excluded, and only blood-associated clones are tabulated. **e**, Distribution of T cells in blood by blood expansion pattern. Bar plots show numbers of T cells found in blood, grouped according to their cluster and further divided by the blood expansion or non-expansion pattern of their parent clone. Because parent clones are guaranteed to have the given cell in blood, blood independence is not possible. **f**, Two-dimensional map of cells in blood by peripheral clonal expansion. Cells from blood with a clonotype are plotted onto the UMAP coordinates from Fig. 2a and coloured green if non-expanded in blood (blood clone size = 1) or a shade of brown for increasing expansion in blood. Ovals from Fig. 2a are added for reference. **g, i**, Distribution of T cell clusters in tumours (**g**) and NAT (**i**) by blood expansion pattern. As in **e**, except for cells in tumour (**g**) and NAT (**i**) from the four patients with blood samples. *P* values are from a chi-square test of counts of cells over T cell clusters in blood versus counts over T cell clusters in tumour (**g**) or NAT (**i**). **h, j**, Two-dimensional maps of cells in tumour (**h**) and NAT (**j**) by blood expansion pattern. As in **f**, except for cells in tumour (**h**) and NAT (**j**) from the four patients with blood samples.



Extended Data Fig. 7 | See next page for caption.

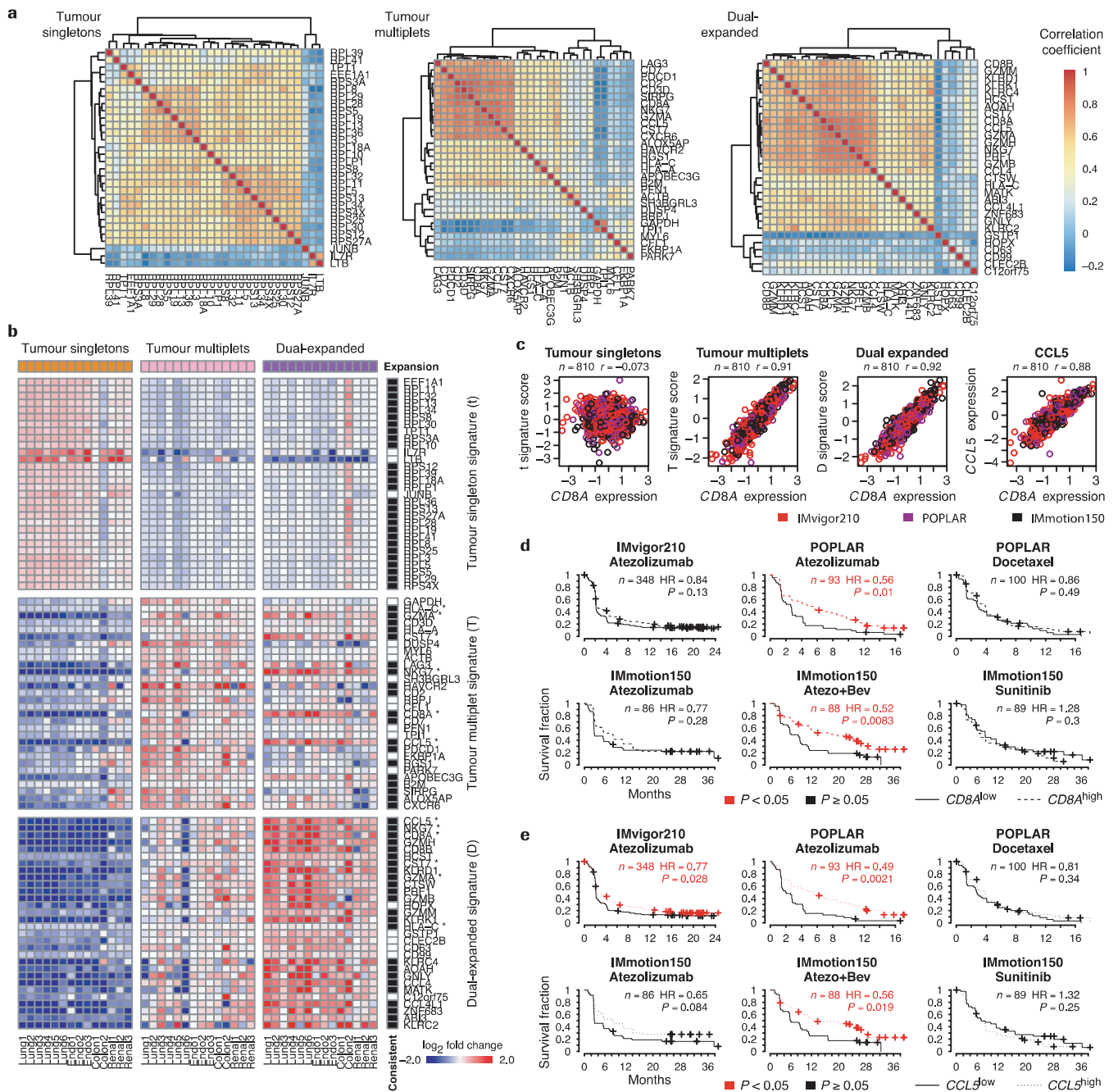
Extended Data Fig. 7 | Sharing of novel clones in tumour with clonotypes in blood. **a**, Matching bulk TCR-seq and scTCR-seq clonotypes. An example from the Yost et al.⁵ dataset is shown to illustrate issues in matching clonotypes across bulk and single-cell technologies. Bulk TCR-seq from Adaptive Biotechnologies immunoSEQ technology yields single 87-base-pair segments of individual β -chains, whereas scTCR-seq from 10x Genomics potentially yields combinations of α - and β -chain CDR3 sequences per clonotype, indicated here by four clonotype IDs and associated sequences in grey boxes. The immunoSEQ output also provides a CDR3 amino acid sequence (bulk-CDR3-aa, rectangle) for productive β -chains, which we used to facilitate matching. We considered clonotypes to match if either β -chain CDR3 from scRNA-seq aligned exactly to the bulk TCR-seq sequence at the nucleotide level, at the position consistent with bulk-CDR3-aa. α -chain CDR3 sequences were disregarded in this process. All four clonotypes shown were therefore considered matches to the bulk TCR-seq sequence. For the purpose of counting T cells, a sum was taken over all matching scRNA-seq clonotypes. Further considerations are provided in Methods. **b**, Correlation of tumour and blood clone sizes in novel CD8 clones. Scatter plots are shown for each patient in Yost et al.⁵ that had both single-cell RNA-seq and TCR-seq of tumour-infiltrating lymphocytes in pre- and post-treatment tumours as well as bulk TCR-seq of T cells in blood. Dots represent novel CD8 clones based on the primary post-treatment cluster from the original analysis. Novel clones are plotted by the count of transcripts in pre-treatment blood (resorting to post-treatment blood for bcc.su002, which lacked a pre-treatment blood sample), used as a proxy for blood clone size, and clone size in post-treatment tumour. Vertical bar

separates novel clones matching a clonotype in blood (blood-associated, right) from those that did not (blood-independent, left). Two-sided P values are shown for a Pearson's correlation coefficient r on blood-associated novel clones. Patients are ordered by their total (blood-associated plus blood-independent) number of novel CD8 clones. Two-sided P values are shown from a Fisher's z -test for the comparison of the correlation coefficient of CD8 novel clones and the correlation coefficient of the CD4 novel clones. **c**, Correlation of tumour and blood clone sizes in novel CD4 clones. Scatter plots are shown for the novel CD4 clones from patients in **a**, in corresponding order, as in **b**. **d**, Clonal diversity in blood. Scatter plots are shown for the patients in **b** and **c**, in corresponding order. Dots represent distinct TCR β -chain rearrangements as provided in the original immunoSEQ analysis, plotted by the numbers of templates reported in pre- and post-treatment blood. For patient bcc.su002, which lacked a pre-treatment blood sample, a one-dimensional strip chart shows the post-treatment TCR repertoire with horizontal jitter added to display points more clearly. Increasing clonal diversity can be observed qualitatively as the increasing presence of clones along the main diagonal, and is quantified using Shannon entropy. **e**, Completeness curves for blood TCR-seq samples. Each plot shows a sample completeness curve for a bulk TCR-seq sample in **d** based on a rarefaction and extrapolation analysis³⁸, with pre- and post-treatment samples coloured as shown. Each curve indicates the estimated coverage of the total set of TCR β -chain rearrangements as a function of the total number of transcripts sampled. Dot indicates the actual number of transcripts sampled, solid lines indicate the interpolated completeness curve, and dashed lines indicate an extrapolation of the completeness curve.



Extended Data Fig. 8 | Matching clonotypes against databases of known and putative virally reactive TCRs. a, TCR repertoire of clonotypes matching VDJdb. A set of rug plots is shown for each patient, with each plot representing the repertoire of clonotypes matching TCRs listed as reacting against common viral antigens from the VDJdb database⁵⁴. Viral antigens shown are from cytomegalovirus (p65 antigen), Epstein–Barr virus (BMLF1, EBNA3) and influenza A (M1). Other antigens from these viruses are listed as ‘other’. Each rug plot depicts each distinct clonotype as a region, coloured by its primary cluster, with the height of each region indicating its total clone size in tumour plus NAT. Clonotypes are stacked on top of one another in random order. In situations in which adjacent clones share the same colour, black lines were used to separate them, when both clones had a clone count greater than 5, indicating a need to resolve them visually. Plots show that most matching clonotypes were singletons, but that patients often had a few virally reactive clonotypes that had expanded greatly. **b**, Association of viral reactivity with

clonal expansion patterns. Clonotypes matching VDJdb and multi-cancer TCRs computed⁵⁵ from The Cancer Genome Atlas (TCGA), suggesting reactivity to a viral antigen, were grouped according to their clonal expansion pattern. Bar plot shows frequencies of matches for each database. *P* values are from a chi-square test on counts of distinct clonotypes. One-sided *P* values are shown next to bars when Bonferroni-adjusted $P < 0.05$ from post hoc Fisher tests performed over the same counts of clonotypes as the chi-square test. **c**, Association of viral reactivity with primary cell clusters. Clonotypes matching VDJdb (left) and multi-cancer TCRs from TCGA (right) were grouped according to their primary cluster. Bar plots show frequencies of matches for each cluster, with bars coloured as in **a**. Vertical bars show the mean fraction of clonotype matches across all clonotypes. *P* values are from a chi-square test on counts of distinct clonotypes with a primary cluster assigned. One-sided *P* values are shown next to bars when Bonferroni-adjusted $P < 0.05$ from post hoc Fisher tests performed over the same counts of clonotypes as the chi-square test.

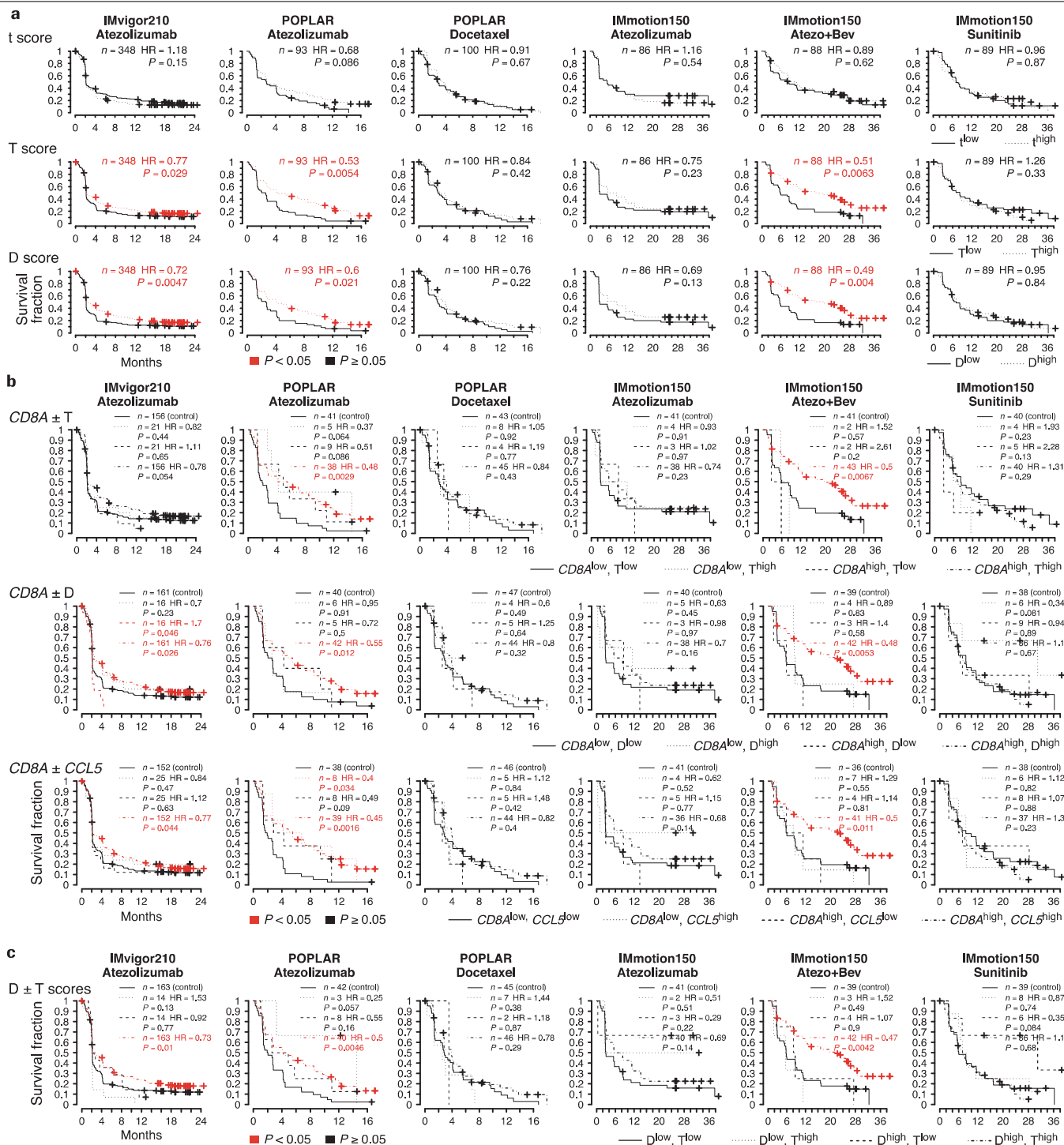


Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | Gene signatures of tissue expansion patterns and relationship to *CD8A* expression. **a**, Consistency of gene signatures in bulk tumour RNA-seq. The 30 highest-ranking genes for each tissue expression pattern involving a tumour sample are shown in a heat map of correlated gene expression from bulk tumour RNA-seq data from all patients in the three clinical trials analysed in this study. Intensities of each cell represent the Pearson's correlation coefficient on the gene expression values from patients between each pair of genes, and dendrograms indicate the hierarchical clustering of genes. The first division of each dendrogram is used to eliminate genes that are inconsistent with other genes in the signature, possibly due to expression by non-T cells. **b**, Expansion signatures in scRNA-seq data. Heat map shows the relative gene expression of signatures from **a** in the scRNA-seq data of this study, used for the initial selection of gene signatures. Intensities indicate the mean \log_2 -transformed fold change of each gene (rows) across patients (columns), in which the fold change was computed within each patient for the T cells from the tumour compartment of a given cluster against all other T cells from the tumour compartment from that patient. 'Consistent' indicates genes that passed the filtering step in **a** by black cells, which are used for subsequent analysis. Genes common to more than one signature are marked by

an asterisk. **c**, Correlation of expansion signatures with *CD8A* expression. Scatter plots show the correlation of *CD8A* expression with expansion signature scores across patient bulk RNA-seq samples from three clinical trials. Expression of *CCL5* is also included as a marker of expansion, ranking highly in both the tumour multiplet and dual expansion signatures from the scRNA-seq analysis. Each dot represents a pre-treatment bulk tumour RNA-seq sample, coloured by the clinical trial. **d**, Survival analysis of *CD8A* expression. Kaplan–Meier plots of PFS are shown for each arm in a clinical trial, with patients dichotomized by their expression of the *CD8A* gene in bulk tumour above (*CD8A*^{high}) or below (*CD8A*^{low}) the median expression among all patients in the corresponding clinical trial. *CD8A* expression is used as a marker for the prevalence of intratumoural CD8⁺ T cells, which is known to be a predictor of response to cancer immunotherapy. Censored observations are indicated by a plus symbol. Hazard ratios (HR) and two-sided *P* values from a Cox proportional-hazards model on patients in both groups are shown, highlighted in red with the associated survival curve when *P* < 0.05. Six patients in IMvigor210 were omitted owing to missing values for PFS. **e**, As in **d**, except for gene expression of *CCL5*.



Extended Data Fig. 10 | Survival analysis. **a**, Survival based on clonal expansion patterns. Kaplan-Meier survival curves for PFS are shown for three clonal expansion signatures (rows) in each arm of a clinical trial (columns), in which patients are dichotomized by scores above (dotted lines) and below (solid lines) the median in each clinical trial. Plus symbols indicate censoring events. Hazard ratios (HR) and two-sided P values from a Cox proportional-hazards model on patients in both groups are shown, highlighted in red with the corresponding survival curve when $P < 0.05$. **b**, Survival based on expansion signatures in the context of $CD8A$ expression. Kaplan-Meier plots for PFS are shown using both $CD8A$ expression and expansion signature scores (rows) in each arm of a clinical trial (columns). Patients in each clinical trial were divided into four groups based on $CD8A$ expression above ($CD8A^{high}$) or below ($CD8A^{low}$)

the median and on expansion signature above (signature^{high}) or below (signature^{low}) the median among all patients in the corresponding clinical trial. Patients with low $CD8A$ expression and low expansion signature score were used as a control for each of the other three groups. Hazard ratios and two-sided P values are from a Cox proportional-hazards model on patients in each group, highlighted in red with the corresponding survival curve when $P < 0.05$. **c**, Survival based on dual-expanded and tumour multiplet signatures. Kaplan-Meier survival curves are plotted as in **b**, except that patients were divided into four groups based on dual-expanded clone signature above (D^{high}) or below (D^{low}) the median and on tumour multiplet signature above (T^{high}) or below (T^{low}) the median among all patients in the corresponding clinical trial.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data were collected using Cell Ranger software (10x Genomics, Pleasanton, CA) version 2.2.0 [Zheng 2017].

Data analysis

Data were analyzed using Perl version 5.18.2, R version 3.6.0, and the following packages and versions in R for analysis: Seurat, 3.0.2 [Stuart 2019]; SingleR, 1.0.1 [Aran 2019]; RankProd, 3.11.0 [Del Carratore 2017]; GSEABase, 1.47.0; limma, 3.41.15 [Ritchie 2015]; annotate_1.63.0; homologene, 1.4.68.19.3.27; mouse4302.db, 3.2.3; cocor, 1.1-3 [Diedenhofen & Musch 2015]; iNEXT, 2.0.19 [Chao 2014]; and survival, 2.44-1.1 [Therneau & Grambsch 2000]. Two-dimensional gene expression maps were generated using coordinates from the Uniform Manifold Approximation and Projection (UMAP) algorithm [McInnes and Healy 2018, Becht 2019], as implemented in the umap-learn module version 0.3.10 in Python version 3.6.0, run under R using reticulate version 1.13. Figures were produced using the following packages and versions in R: colorspace, 1.4-1 [Zeileis 2019]; RColorBrewer, 1.1-2; pheatmap, 1.0.12; and superheat, 0.1.0 [Barter & Yu 2018]. External datasets were obtained using GenomicDataCommons, 1.9.0 [Morgan & Davis 2017] and GEOquery, 2.53.0 [Davis & Meltzer 2007].

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

FASTQ files containing raw reads from the scRNA-seq and scTCR-seq analyses have been deposited with the European Genome-phenome Archive (GA) under studies EGAS00001003993 and EGAS00001003994, and datasets EGAD00001005464 and EGAD00001005465. These files are available under controlled access upon request to the Data Access Committee, with contact information provided at the corresponding Web page at EGA. Processed output files from Cell Ranger, integrated assay results from Seurat, and metadata with UMAP coordinates, cluster assignments, and clonotypes are available from the NCBI Gene Expression Omnibus database under accession GSE139555, which also provides computer code used to generate the analyses and figures in this paper as a Supplementary File.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size of patients was determined in part by the exploratory nature of our study, the novelty of the 10x Genomics single-cell TCR-seq assay technology, and the relative difficulty of obtaining surgical samples. However, our study of 14 patients is comparable to published studies of single-cell RNA-seq and single-cell TCR-seq data, with 2018 studies having 12–14 patients using another technology (SmartSeq2). Furthermore, our dataset has greater sequencing depth than published studies, providing evidence that our sample sizes of cells and clones should be adequate for comparable analysis of T cell subtypes and clonal expansion. Sample sizes for clinical trial analyses were determined by the designers of the clinical trials.
Data exclusions	Single cells from CD3- and CD45-selected samples were separated computationally into T and non-T cells, as described in Methods and documented in Extended Data Fig 1a–g. Patient bcc.su003 from Yost et al., 2019 showed no overlapping clones between pre- and post-treatment tumour, and could not be analyzed further, as described in Methods.
Replication	We tested our script for identifying clonotypes against the assignment from CellRanger software for each patient. We compared our cluster analysis of T cell subsets against the clusters obtained by Guo et al., 2018, Zhang et al., 2018, and Yost et al., 2019. We confirmed our relationship between peripheral clonal expansion and infiltration, as well as the correlation between blood clone size and tumor clone size, using 14 patients from Guo et al., 2018 and 12 patients from Zhang et al., 2018. We observed relationships between T cell subsets and clonal expansion patterns similar to ours using the same two datasets. We observed a correlation between blood transcript counts and novel tumour clone size on two patients in Yost et al., 2019 with bulk RNA-seq measurements in blood and the largest numbers of novel clones. We confirmed our hypotheses regarding the replenishment of T cells from blood using 6 patients from Yost et al., 2019. We confirmed our analysis of viral reactivity using two datasets: VDJdb and TCGA-derived TCR sequences. We performed survival analysis on gene signatures for tissue expansion patterns using three independent clinical trials.
Randomization	Single-cell TCR data were grouped using T cell receptor chains and CDR3 nucleotide sequences assigned by the Cell Ranger software package. Single-cell RNA-Seq data were grouped using existing classification algorithms from the Seurat package. All two-dimensional gene expression maps plotted cells in random order. Random jitter was added to scatterplots to visually display all points.
Blinding	Patient samples were obtained from commercial vendors, who were blinded to the scientific purpose of the samples. Sample preparation was performed by laboratory personnel who were blinded to the subsequent computational analysis. Exploratory computational analyses were performed in an unsupervised manner, with analyses blinded to preconceived notions of the expected outcome. Statistical tests were performed without bias about expected results. Clinical data were obtained from randomized clinical trials that blinded both patients and practitioners about their treatment arms.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials Samples no longer available, because they were consumed in the preparation for analysis

Antibodies

Antibodies used	EpCAM: Biolegend 9C4 antibody, Cat# 324208, Lot# B198775, fluorochrome APC. CD45: Biolegend 2D1 antibody, Cat# 368516, Lot# B264243, fluorochrome APC/Cy7. CD3: BD Biosciences SK7 antibody, Cat# 564001, Lot# 6109971, fluorochrome BUV395.
Validation	EpCAM (Biolegend 9C4 antibody): Product information: https://www.biolegend.com/en-us/search-results/apc-anti-human-cd326-epcam-antibody-3758 CD45 (Biolegend 2D1 antibody): https://www.biolegend.com/en-us/search-results/apc-cy7-anti-human-cd45-antibody-12400 CD3 (BD Biosciences SK7 antibody): https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/surface-markers/human/buv395-mouse-anti-human-cd3-sk7-also-known-as-leu-4/p/564001 Certificates of analysis for Biolegend lots are available at http://biolegend.com/it-it/certificate-of-analysis . A quality certificate for the BD Biosciences lot is available at http://regdocs.bd.com/regdocs/searchCOAAAction.do .

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The 14 patients studied by single-cell TCR and scRNA-seq are as follows: Lung1: age 65, female. Lung2: age 60, female. Lung3: age 51, female. Lung4: age 67, female. Lung5: age 71, female. Lung6: age 55, female. Endo1: age 59, female. Endo2: age 62, female. Endo3: age 61, female. Colon1: age 47, female. Colon2: age 80, female. Renal1: age 65, male. Renal2: age 45, female. Renal3: age 62, male. Population characteristics of clinical trial subjects are presented in their corresponding publications.
Recruitment	Patient samples were procured from third parties (Discovery Life Sciences, iSpecimen Inc., Avaden BioSciences, and TriMetis Life Sciences) from adult patients undergoing resection surgery.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Fresh surgical samples were separated into tumour and NAT compartments by the reviewing pathologist and shipped overnight to our institution. Upon arrival, samples were rinsed with phosphate-buffered saline (PBS) until no traces of blood were visually detected. Subsequently, samples were digested with a combination of Collagenase D (0.5mg/ml) and DNase (0.1mg/ml) for 15min at 37°C with gentle shaking. Subsequently, samples were subjected to a gentleMACS dissociator (Miltenyi Biotec), followed by an additional 10 min incubation at 37°C. Peripheral blood mononuclear cells (PBMCs) from patient blood samples were isolated using 50mL Leucosep™ tubes (Greiner Bio-One International, Germany) and Ficoll-Paque™ PLUS (GE Healthcare, Sweden). Whole blood drawn into sodium heparin blood collection tubes were diluted 3 times with PBS without calcium or magnesium (Lonza, Walkersville, MD). Diluted cell suspensions were carefully layered on Leucosep tubes and centrifuged for 15 minutes at 800 x g at room temperature (RT). Interphase-containing PBMCs were harvested and washed with PBS and subsequently centrifuged for 10 minutes at 250 x g at RT before further processing.
Instrument	Cells were purified by fluorescence-activated cell sorting (FACS) on a Becton Dickinson FACSria Fusion cell sorter equipped with

Instrument	4 lasers (405nm, 488nm, 561nm, 638nm). A 70-micron nozzle running at 70 psi and 90kHz was used as the setup for each sort session.
Software	FACSDiva (version 8.0.1) and FlowJo (version 10) were used to collect and analyze the flow cytometry data.
Cell population abundance	Post-sort samples were analyzed for purity using the same FACS sorter. Using the same gating strategy, the purity of the samples was determined and calculated to be 97–99%
Gating strategy	Before gating on fluorescence, live, single cells were gated using FSC-A and SSC-A (for intact cells) and SSC-W/SSC-H and FSC-W/FSC-H (to ensure that only singlets were sorted). FACS gates were drawn to include only live single cells based on Calcein Blue AM+ and 7-AAD (Thermo Fisher Scientific). Further gates were drawn to arrive at CD3+CD45+EpCAM- (for CD3+ selected samples) or CD45+EpCAM- cells (for CD45+ selected samples). Boundaries between positive and negative cell fractions were drawn based on single-color stains.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Glucagon stimulates gluconeogenesis by INSP3R1-mediated hepatic lipolysis

<https://doi.org/10.1038/s41586-020-2074-6>

Received: 24 February 2018

Accepted: 15 January 2020

Published online: 4 March 2020

 Check for updates

Rachel J. Perry^{1,2}, Dongyan Zhang¹, Mateus T. Guerra¹, Allison L. Brill², Leigh Goedeke¹, Ali R. Nasiri¹, Aviva Rabin-Court¹, Yongliang Wang¹, Liang Peng¹, Sylvie Dufour¹, Ye Zhang¹, Xian-Man Zhang¹, Gina M. Butrico¹, Keshia Toussaint¹, Yuichi Nozaki¹, Gary W. Cline¹, Kitt Falk Petersen¹, Michael H. Nathanson¹, Barbara E. Ehrlich^{2,3} & Gerald I. Shulman^{1,2✉}

Although it is well-established that reductions in the ratio of insulin to glucagon in the portal vein have a major role in the dysregulation of hepatic glucose metabolism in type-2 diabetes^{1–3}, the mechanisms by which glucagon affects hepatic glucose production and mitochondrial oxidation are poorly understood. Here we show that glucagon stimulates hepatic gluconeogenesis by increasing the activity of hepatic adipose triglyceride lipase, intrahepatic lipolysis, hepatic acetyl-CoA content and pyruvate carboxylase flux, while also increasing mitochondrial fat oxidation—all of which are mediated by stimulation of the inositol triphosphate receptor 1 (INSP3R1). In rats and mice, chronic physiological increases in plasma glucagon concentrations increased mitochondrial oxidation of fat in the liver and reversed diet-induced hepatic steatosis and insulin resistance. However, these effects of chronic glucagon treatment—reversing hepatic steatosis and glucose intolerance—were abrogated in *Insp3r1* (also known as *Itpr1*)-knockout mice. These results provide insights into glucagon biology and suggest that INSP3R1 may represent a target for therapies that aim to reverse nonalcoholic fatty liver disease and type-2 diabetes.

On the basis of the clear role for glucagon in the pathogenesis of diabetes, glucagon-blocking therapies have been pursued as adjuncts to therapies for this condition. Glucagon antagonism^{4–9} and knock-down of the glucagon receptor^{10,11} have shown glucose-lowering effects; however, these agents increase liver enzymes^{7,8} by an unknown mechanism. Conversely, a dual agonist for glucagon-like peptide-1 and glucagon receptor increases energy expenditure and promotes weight loss, an effect associated with reductions in blood glucose¹². Together, these data suggest a role for glucagon in promoting hepatic mitochondrial fat oxidation, in addition to its known effect of stimulating hepatic gluconeogenesis.

Glucagon stimulates hepatic glycogenolysis and gluconeogenesis, the latter of which is believed to occur largely through transcriptional regulation. Hepatic calcium signalling is integral to the transcriptional regulation of gluconeogenesis: inhibition or deletion of liver Ca²⁺/calmodulin-dependent protein kinase II (CAMKII) results in reduced expression of proteins in the gluconeogenic pathway in mouse livers and associated reductions in plasma glucose and insulin concentrations^{13–15}. Type-1 INSP3R is the isoform that is primarily responsible for mitochondrial calcium signalling in hepatocytes¹⁶. It has previously been shown¹⁴ that knocking down *Insp3r1* reduces glucose production in isolated hepatocytes; however, no studies have fully characterized the role and mechanism of INSP3R signalling in glucagon-stimulated hepatic gluconeogenesis in vivo.

To examine the potential calcium dependence of the acute response to glucagon, we studied mice with *Insp3r1* knocked out specifically in the liver, along with their weight-matched wild-type littermates

(Extended Data Fig. 1a). The *Insp3r1*-knockout mice exhibited a marked reduction in hepatic INSP3R1 protein expression (Extended Data Fig. 1b; full gels for all blots are shown in Supplementary Fig. 1). We examined subcellular fractions in primary hepatocytes and found that INSP3R1 was associated with mitochondria, whereas the type-2 INSP3R isoform (INSP3R2) was not; the type-3 INSP3R isoform (INSP3R3) was not detected at all (Extended Data Fig. 1c), consistent with previous data that show that INSP3R1 is primarily responsible for mitochondrial calcium signals¹⁶. We then treated mice—which were fasted overnight, and therefore depleted of glycogen—with an acute infusion of glucagon, which modestly increased plasma glucose and insulin concentrations in wild-type—but not *Insp3r1*-knockout—mice and increased phosphorylation of INSP3R1 (Fig. 1a–c, Extended Data Fig. 1c). Phosphorylation of CREB-regulated transcriptional coactivator 2 (CRTC2) was decreased with glucagon infusion in both wild-type and *Insp3r1*-knockout mice (Extended Data Fig. 1d), consistent with previous literature¹⁴. Glucagon increased cAMP concentrations and the activity of protein kinase A (PKA) in both wild-type and *Insp3r1*-knockout mice, dissociating these mediators from glucagon-induced changes in plasma glucose concentrations (Fig. 1d, e). Consistent with a previous report¹³, glucagon stimulated the phosphorylation of CAMKII but not of CAMKIV in an INSP3R1-dependent manner (Fig. 1f, Extended Data Fig. 1e). These data imply that INSP3R1 activation and CAMKII activity are required for the acute effect of glucagon in stimulating hepatic glucose production (HGP), which can be attributed almost entirely to gluconeogenesis in this hepatic glycogen-depleted state¹⁷. Indeed, we measured increases of approximately 25% in in vivo HGP, which could be attributed to an

¹Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. ²Department of Cellular and Molecular Physiology, Yale School of Medicine, New Haven, CT, USA. ³Department of Pharmacology, Yale School of Medicine, New Haven, CT, USA. ✉e-mail: gerald.shulman@yale.edu

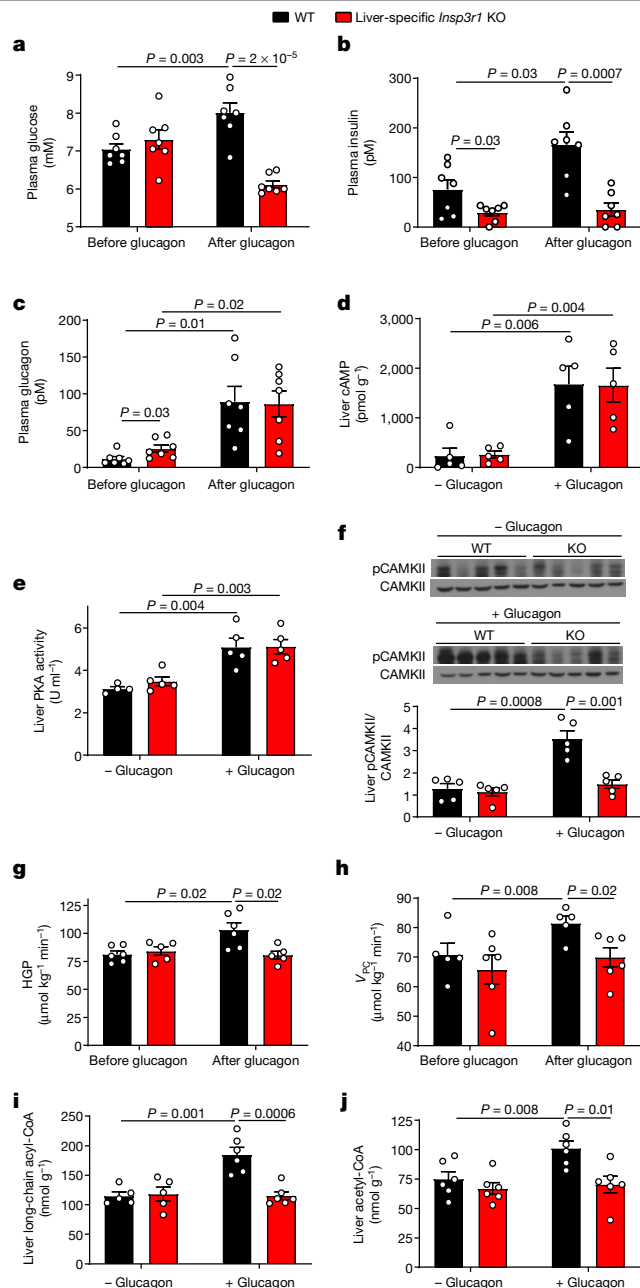


Fig. 1 | Glucagon acutely stimulates hepatic gluconeogenesis by increasing hepatic acetyl-CoA content and V_{PC} . **a–c**, Plasma glucose, insulin and glucagon concentrations before and at the end of a 2-h intravenous infusion of glucagon ($n = 7$). **d–f**, Hepatic cAMP concentrations, PKA activity and CAMKII phosphorylation (pCAMKII) ($n = 5$, with the exception of wild type without (–) glucagon in **e** ($n = 4$)). Blots in **f**, and in Fig. 2a, Extended Data Figs. 1b, c, e, 3f, g, 4a were stripped and reprobed for all proteins of interest. **g, h**, HGP ($n = 6$ wild type and 5 knockout) and V_{PC} ($n = 5$ wild type and 6 knockout). **i, j**, Hepatic long-chain acyl- ($n = 5$ – glucagon, 6 with (+) glucagon) and acetyl-CoA content ($n = 6$). In all panels, mean \pm s.e.m. is shown. Groups were compared before and after glucagon treatment (**a–c, g** and **h**) by two-tailed paired Student's *t*-test, and separate mice (\pm glucagon in **d–f, i, j**, and wild-type versus knockout mice in all panels) were compared by two-tailed unpaired Student's *t*-test. All *n* values refer to numbers of mice.

increased rate of hepatic pyruvate carboxylase flux (V_{PC}) that was associated with amino acid depletion in the liver without alterations in hepatic glycogen content (Fig. 1g, h, Extended Data Fig. 1f–m, Supplementary Tables 1, 2). The in vivo HGP data were mirrored by twofold increases in both HGP and V_{PC} in isolated hepatocytes from wild-type

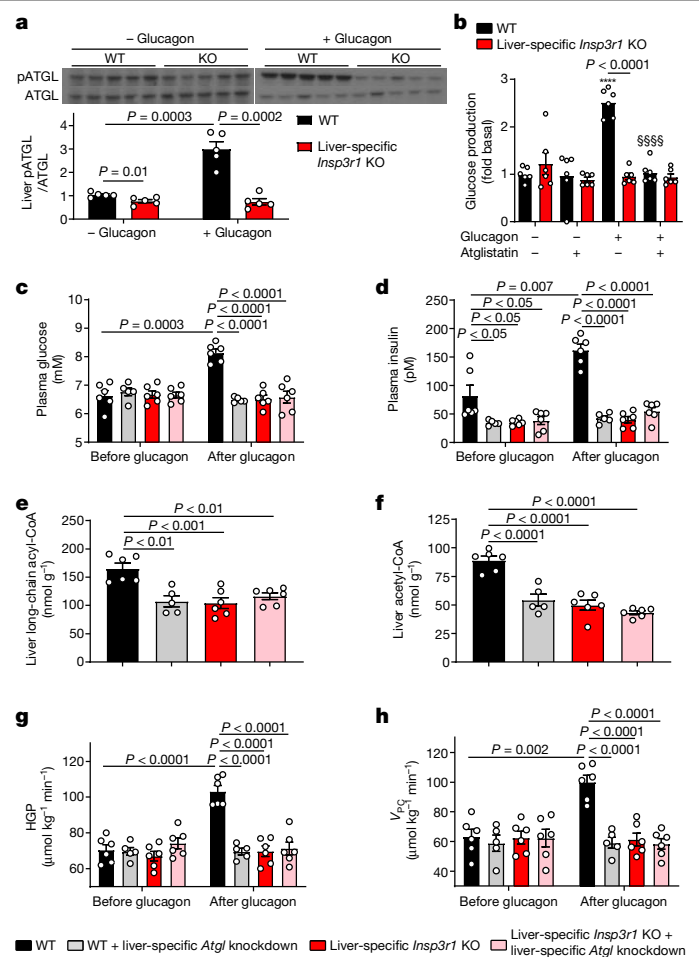


Fig. 2 | Glucagon requires INSP3-mediated intrahepatic lipolysis to promote V_{PC} and HGP. **a**, Phosphorylation of ATGL at Ser406 ($n = 5$). The blots here, as well as those in Fig. 1f, Extended Data Figs. 1b, c, e, 5f, g, 7a, were stripped and reprobed for all proteins of interest. **b**, Glucose production in hepatocytes ($n = 6$ mice per group) incubated with the ATGL inhibitor atglstatin and/or glucagon. **** $P < 0.0001$ versus wild type – glucagon – atglstatin, \$\$\$\$ $P < 0.0001$ versus wild type + glucagon – atglstatin. **c, d**, Plasma glucose and insulin concentrations in mice treated with an adeno-associated virus to knock down *Atgl* in a liver-specific manner. In **c–h**, $n = 6$ wild type, 5 wild type + *Atgl* knockdown and 6 knockout \pm *Atgl* knockdown. **e, f**, Liver long-chain acyl- and acetyl-CoA concentrations following a 2-h glucagon infusion. **g, h**, HGP and V_{PC} . In all panels, mean \pm s.e.m. is shown. Groups were compared before and after glucagon treatment (**c, d**) by two-tailed paired Student's *t*-test, and groups in all other panels (as well as the four separate groups in **c, d**) were compared by two-tailed unpaired Student's *t*-test. All *n* values refer to numbers of mice.

but not *Insp3r1*-knockout mice incubated with glucagon (Extended Data Fig. 1n, o). The ability of insulin to suppress HGP and V_{PC} was unaltered in *Insp3r1*-knockout hepatocytes, and malic enzyme activity was dissociated from glucose production by treatment with a malic enzyme inhibitor (Extended Data Fig. 1p–s). Finally, in vivo studies with somatostatin and replacement insulin infusion to mimic basal insulin concentrations in the portal vein (85–90 pM in both genotypes) revealed that changes in plasma insulin could not explain the observation that glucagon stimulated HGP and V_{PC} only in wild-type mice (Extended Data Fig. 1t–x).

To further explore the calcium-dependent mechanism by which glucagon stimulates HGP and V_{PC} , we treated hepatocytes with two phospholipase C (PLC) antagonists (ET-18-OCH₃ or U-73122) and found that both agents inhibited glucose production and V_{PC} in wild-type – but not *Insp3r1*-knockout – hepatocytes (Extended Data Fig. 2a–d). Similarly, a PKA inhibitor (H-89) abrogated the ability of glucagon to

stimulate glucose production and V_{PC} only in wild-type hepatocytes, which demonstrates that both the PLC and PKA pathways are required to activate gluconeogenesis in response to glucagon via INSP3R1 signalling (Extended Data Fig. 2e, f). To specifically confirm the role of INSP3R1 in promoting HGP, we treated hepatocytes with vasopressin (an activator of the INSP3 receptor) and found that this agent recapitulated the effect of glucagon to increase glucose production and V_{PC} (Extended Data Fig. 2g, h). However, small-molecule inhibitors of INSP3 (2-aminoethoxydiphenyl borate (2-APB) and caffeine), CAMKII and CAMKIV (KN-93) and the SERCA pump (thapsigargin) inhibited the ability of glucagon to stimulate glucose production and V_{PC} (Extended Data Fig. 2i–p), which suggests that maintaining a normal balance of calcium throughout the cell is required for normal INSP3R1 signalling. None of these agents had any effect in *Insp3r1*-knockout mice, which indicates that INSP3R2 and INSP3R3 do not restore glucose production in the absence of INSP3R1 in the liver.

The observed increases in the rate of HGP occurred in the absence of any effect of glucagon to acutely increase the expression of mRNAs or proteins associated with hepatic gluconeogenesis in vivo. However, glucagon infusion increased hepatic long-chain acyl-CoA and acetyl-CoA content by 35–60% (Fig. 1i, j, Extended Data Fig. 3a–e). These increases in hepatic long-chain acyl- and acetyl-CoA concentrations were dissociated from the phosphorylation of both acetyl-CoA carboxylase (ACC) and 5' AMP-activated protein kinase (AMPK), both of which were increased by glucagon in *Insp3r1*-knockout, but not in wild-type, mice (Extended Data Fig. 3f, g). The differences in ACC and AMPK phosphorylation between genotypes probably reflect the interplay between the ability of glucagon to promote the phosphorylation of both of these enzymes and the ability of hyperinsulinaemia (which was observed only in wild-type mice infused with glucagon) to suppress this phosphorylation. Hepatic long-chain acyl-CoA and acetyl-CoA content increased in wild-type mice infused with glucagon despite reduced plasma non-esterified fatty acids (NEFA) concentrations (Extended Data Fig. 3h), consistent with glucagon stimulation of intrahepatic lipolysis but not white adipose tissue lipolysis. As acetyl-CoA is an allosteric activator of pyruvate carboxylase (PC)¹⁸, the increase in hepatic acetyl-CoA content observed with glucagon infusion could explain the increases in V_{PC} and HGP that occurred after glucagon treatment in wild-type mice. These changes occurred independently of any changes in hepatic malonyl-CoA content (Extended Data Fig. 3i), which indicates that malonyl-CoA suppression of carnitine palmitoyl transferase I was not responsible for glucagon stimulation of hepatic mitochondrial β -oxidation under these conditions.

To further understand the physiological role of glucagon-induced HGP, we fasted wild-type and *Insp3r1*-knockout mice for 48 h. In the starved state and despite 70–90% increases in plasma glucagon concentrations in the tail vein and portal vein, *Insp3r1*-knockout mice manifested lower plasma glucose and insulin concentrations that were associated with reductions in hepatic long-chain acyl-CoA and acetyl-CoA content and no change in hepatic malonyl-CoA content as compared to wild-type mice (Extended Data Fig. 3j–o). These data demonstrate a critical role for INSP3R1 in the maintenance of euglycaemia during starvation.

We next hypothesized that INSP3R1 stimulation of intrahepatic lipolysis may explain the observed increases in hepatic acetyl-CoA content and V_{PC} after glucagon treatment. Consistent with this hypothesis, phosphorylation of hepatic adipose triglyceride lipase (ATGL) at Ser406—which has previously been shown to regulate the activity of ATGL¹⁹—was increased threefold by glucagon in wild-type, but not *Insp3r1*-knockout, mouse livers (Fig. 2a). By contrast, phosphorylation of hormone-sensitive lipase (HSL) was increased by glucagon in both genotypes, which dissociates HSL activity from glucagon activation of HGP and V_{PC} (Extended Data Fig. 4a). In vitro studies revealed 60–100% increases in NEFA and glycerol production with glucagon treatment in wild-type, but not INSP3R1-deficient, hepatocytes (Extended

Data Fig. 4b, c). Plasma NEFA concentrations were unchanged by glucagon infusion in mice treated with somatostatin (Extended Data Fig. 4d). These data suggest that although glucagon may promote lipolysis when infused at markedly supraphysiological doses²⁰, in the setting of physiological concentrations of glucagon and intact β -cell function, glucagon does not directly affect white adipose tissue lipolysis^{21–23}. Confirming the requirement for glucagon-stimulated intrahepatic lipolysis to promote gluconeogenesis, glucagon had no effect in hepatocytes treated with a small molecule inhibitor of ATGL (atglistatin). Similarly, INSP3 agonism recapitulated the effect of glucagon in terms of stimulating intrahepatic lipolysis, and inhibitors of INSP3, CAMKII and CAMKIV, PLC, PKA and the SERCA pump inhibited the ability of glucagon to stimulate intrahepatic lipolysis, glucose production and V_{PC} (Fig. 2b, Extended Data Fig. 4e–n). To examine the effect of glucagon stimulation of intrahepatic lipolysis in vivo, we knocked down *Atgl* (also known as *Pnpla2*) in a liver-specific manner in wild-type mice and their *Insp3r1*-knockout littermates (Extended Data Fig. 5a–c). *Atgl* knockdown abrogated the ability of glucagon to stimulate HGP, V_{PC} and to increase long-chain acyl- and acetyl-CoA content in wild-type mice (Fig. 2c–h), which demonstrates a critical role for the stimulation of intrahepatic lipolysis in mediating the increases in each parameter that result from glucagon infusion. These alterations in hepatic gluconeogenesis were again dissociated from changes in hepatic glycogen content or the expression of gluconeogenic proteins in the liver. In addition, the alterations in the rate of hepatic gluconeogenesis were dissociated from white adipose tissue lipolysis and from hepatic malonyl-CoA content (Extended Data Fig. 5d–j). Taken together, these data demonstrate that glucagon acutely stimulates hepatic gluconeogenesis via INSP3R1 by promoting intrahepatic lipolysis through the stimulation of ATGL, and thereby increases hepatic acetyl-CoA and gluconeogenesis via allosteric activation of PC.

We hypothesized that glucagon may also stimulate hepatic mitochondrial oxidation via INSP3R1. Glucagon increased both mitochondrial and cytosolic calcium signalling in hepatocytes from wild-type mice, whereas *Insp3r1*-knockout mice manifested a reduced response to glucagon, with some residual calcium responsiveness that was probably attributable to the activity of INSP3R2 and INSP3R3. Vasopressin caused a similar increase in cytosolic and mitochondrial calcium signalling in wild-type, but not *Insp3r1*-knockout, mice. However, incubation of wild-type hepatocytes with a PLC inhibitor completely abrogated the mitochondrial and cytosolic calcium responses to glucagon. PKA inhibition reduced the amplitude of the mitochondrial—but not the cytosolic—calcium response, and lowered the percentage of responding cells by 75% (Extended Data Fig. 6a–m). Ex vivo positional isotopomer nuclear magnetic resonance (NMR) tracer analysis (PINTA) revealed that glucagon stimulated hepatic mitochondrial oxidation (measured as the rate of citrate synthase flux (V_{CS})) in vivo, increasing V_{CS} fivefold in wild-type mice; this could mostly be attributed to increased mitochondrial fat oxidation (Fig. 3a, b). However, these increases in hepatic mitochondrial oxidation in response to glucagon did not occur in *Insp3r1*-knockout mice. Taken together, these data suggest that glucagon stimulates hepatic mitochondrial oxidation through the activation of INSP3R1, which results in increased levels of intramitochondrial calcium and—in turn—stimulation of mitochondrial dehydrogenases²⁴. Consistent with this hypothesis, we observed a sevenfold increase in the rate of hepatic pyruvate dehydrogenase flux (V_{PDH}) in wild-type, but not *Insp3r1*-knockout, mice infused with glucagon, without any difference in the rate of pyruvate kinase flux (V_{PK}) (Extended Data Fig. 6n, o). Glucagon stimulation of mitochondrial oxidation was not dependent on alterations in insulin: wild-type mice infused with somatostatin, basal insulin and approximately 100-pM glucagon exhibited increases in V_{CS} similar to those of wild-type mice not treated with somatostatin—however, again, no glucagon stimulation of V_{CS} was observed in their *Insp3r1*-knockout littermates (Extended Data Fig. 6p). In vitro studies confirmed an increase in oxygen consumption with glucagon

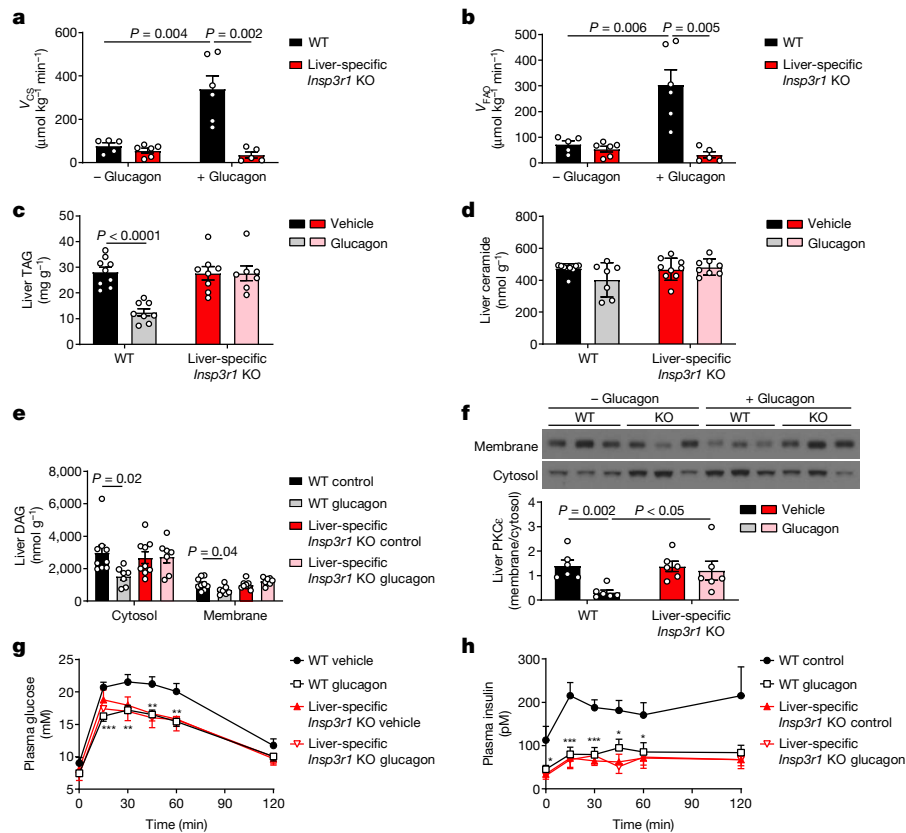


Fig. 3 | Chronic increases in mitochondrial oxidation with a continuous 3.5-week glucagon infusion reverse hepatic steatosis and improve glucose tolerance in an INSP3R1-dependent manner. **a, b**, Liver V_{CS} and the rate of fatty acid oxidation (V_{FAO}) with acute glucagon infusion. $n = 5$ wild type – glucagon, 6 knockout – glucagon, 6 wild type + glucagon and 5 knockout + glucagon. **c**, Liver triacylglycerol (TAG) concentrations in high-fat-diet-fed mice chronically infused with glucagon. $n = 9$ wild type – glucagon, 8 wild type + glucagon, 8 knockout – glucagon and 7 knockout + glucagon. **d**, Liver ceramide. $n = 9$ wild

type – glucagon, 7 wild type + glucagon, 9 knockout – glucagon and 8 knockout + glucagon. **e**, Liver diacylglycerol (DAG). $n = 9$ without glucagon, 7 with glucagon. **f**, PKC ϵ translocation ($n = 6$). **g, h**, Plasma glucose and insulin during a glucose tolerance test. $n = 10$ wild type without glucagon (WT vehicle), 11 wild type with glucagon (WT glucagon), 8 knockout (KO) vehicle, 8 knockout glucagon). In all panels, mean \pm s.e.m. is shown. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Groups were compared using two-tailed unpaired Student's *t*-test. All *n* values refer to numbers of mice.

stimulation in isolated hepatocytes from wild-type, but not *Insp3r1*-knockout, mice (Extended Data Fig. 6q). These data demonstrate that glucagon mediates these effects in a cell-autonomous manner.

Liver triglyceride concentrations did not differ between chow-fed wild-type and *Insp3r1*-knockout mice (Extended Data Fig. 6r), which suggests that basal physiological levels of glucagon may not increase mitochondrial fat oxidation in ad libitum fed mice to a sufficient extent to alter hepatic triglyceride content. However, we hypothesized that the chronic increases in hepatic mitochondrial oxidation induced by chronic glucagon treatment would reverse nonalcoholic fatty liver disease and improve whole-body insulin sensitivity. To investigate this, we performed chronic glucagon infusions in awake rats with diet-induced obesity, and found that after ten days of treatment chronic glucagon infusion doubled the rate of hepatic mitochondrial oxidation (which could be attributed to increased rates of hepatic fat oxidation) (Extended Data Fig. 7a–g). Two hours after the withdrawal of glucagon, rats exhibited lower fasting plasma glucose and insulin concentrations that were associated with 50–90% reductions in hepatic triglyceride and diacylglycerol concentrations as well as marked reductions in protein kinase C- ϵ (PKC ϵ) translocation—despite similar food intake, body weight and hepatic ceramide content (Extended Data Fig. 7h–o, Supplementary Tables 3, 4). Consistent with chronic increases in hepatic glycogenolysis, chronic glucagon infusion also resulted in reductions in liver glycogen content. In contrast to acute glucagon infusion, chronic glucagon infusion with withdrawal of glucagon treatment four hours before the tissues were isolated lowered both the hepatic acetyl- and

malonyl-CoA content (Extended Data Fig. 7p–r); the reduction in malonyl-CoA content potentially contributes to the observed reductions in hepatic lipid concentrations due to suppression of hepatic lipogenesis²⁵. Consistent with their lower hepatic lipid and acetyl-CoA content, rats that were chronically infused with glucagon manifested improved glucose tolerance and insulin sensitivity (Extended Data Fig. 7s–v).

To determine whether INSP3R1-dependent calcium signalling mediates the ability of chronic hyperglucagonaemia to reverse nonalcoholic fatty liver disease, we performed a four-week continuous infusion of glucagon in wild-type and *Insp3r1*-knockout mice fed a high-fat diet. Despite unchanged body weight, food and water intake, and energetics, wild-type mice treated with glucagon exhibited 50–80% reductions in hepatic triglyceride and diacylglycerol (but not ceramide) content and in PKC ϵ translocation that resulted in a marked improvement in glucose tolerance, whereas each of these parameters was unchanged in *Insp3r1*-knockout mice (Fig. 3c–h, Extended Data Fig. 8a–j, Supplementary Tables 5, 6).

Collectively, our studies reveal that glucagon stimulates intrahepatic lipolysis through an INSP3R1- and CAMKII-dependent process that increases hepatic acetyl-CoA content (which allosterically activates V_{PC}^{18}), and that this phenomenon explains the acute transcription-independent ability of glucagon to acutely stimulate gluconeogenesis in vivo. In addition, glucagon stimulates hepatic mitochondrial oxidation through INSP3R1-mediated calcium signalling, and this process can be exploited to reverse nonalcoholic fatty liver disease and hepatic insulin resistance with short-term continuous glucagon treatment.

These data reveal a transcription-independent alternative mechanism (Extended Data Fig. 9) for glucagon action. Our studies provide evidence in support of INSP3R1 as a potential target for the treatment of nonalcoholic fatty liver disease, nonalcoholic steatohepatitis and type-2 diabetes; future clinical studies will be required to examine this possibility.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2074-6>.

1. Unger, R. H. Glucoregulatory hormones in health and disease. A teleologic model. *Diabetes* **15**, 500–506 (1966).
2. Unger, R. H. Pancreatic glucagon in health and disease. *Adv. Intern. Med.* **17**, 265–288 (1971).
3. Müller, T. D., Finan, B., Clemmensen, C., DiMarchi, R. D. & Tschöp, M. H. The new biology and pharmacology of glucagon. *Physiol. Rev.* **97**, 721–766 (2017).
4. Brand, C. L. et al. Immunoneutralization of endogenous glucagon with monoclonal glucagon antibody normalizes hyperglycaemia in moderately streptozotocin-diabetic rats. *Diabetologia* **37**, 985–993 (1994).
5. Sørensen, H. et al. Immunoneutralization of endogenous glucagon reduces hepatic glucose output and improves long-term glycemic control in diabetic *ob/ob* mice. *Diabetes* **55**, 2843–2848 (2006).
6. Okamoto, H. et al. Glucagon receptor inhibition normalizes blood glucose in severe insulin-resistant mice. *Proc. Natl Acad. Sci. USA* **114**, 2753–2758 (2017).
7. Kazda, C. M. et al. Evaluation of efficacy and safety of the glucagon receptor antagonist LY2409021 in patients with type 2 diabetes: 12- and 24-week phase 2 studies. *Diabetes Care* **39**, 1241–1249 (2016).
8. Guzman, C. B. et al. Treatment with LY2409021, a glucagon receptor antagonist, increases liver fat in patients with type 2 diabetes. *Diabetes Obes. Metab.* **19**, 1521–1528 (2017).
9. Kazda, C. M. et al. Treatment with the glucagon receptor antagonist LY2409021 increases ambulatory blood pressure in patients with type 2 diabetes. *Diabetes Obes. Metab.* **19**, 1071–1077 (2017).
10. Liang, Y. et al. Reduction in glucagon receptor expression by an antisense oligonucleotide ameliorates diabetic syndrome in *db/db* mice. *Diabetes* **53**, 410–417 (2004).
11. Sloop, K. W. et al. Hepatic and glucagon-like peptide-1-mediated reversal of diabetes by glucagon receptor antisense oligonucleotide inhibitors. *J. Clin. Invest.* **113**, 1571–1581 (2004).
12. Henderson, S. J. et al. Robust anti-obesity and metabolic effects of a dual GLP-1/glucagon receptor peptide agonist in rodents and non-human primates. *Diabetes Obes. Metab.* **18**, 1176–1190 (2016).
13. Ozcan, L. et al. Calcium signaling through CaMKII regulates hepatic glucose production in fasting and obesity. *Cell Metab.* **15**, 739–751 (2012).
14. Wang, Y. et al. Inositol-1,4,5-trisphosphate receptor regulates hepatic gluconeogenesis in fasting and diabetes. *Nature* **485**, 128–132 (2012).
15. Ozcan, L. et al. Activation of calcium/calmodulin-dependent protein kinase II in obesity mediates suppression of hepatic insulin signaling. *Cell Metab.* **18**, 803–815 (2013).
16. Ferioli, C. N. et al. Hepatic inositol 1,4,5 trisphosphate receptor type 1 mediates fatty liver. *Hepatol. Commun.* **1**, 23–35 (2017).
17. Perry, R. J. et al. Leptin mediates a glucose-fatty acid cycle to maintain glucose homeostasis in starvation. *Cell* **172**, 234–248 (2018).
18. Perry, R. J. et al. Hepatic acetyl CoA links adipose tissue inflammation to hepatic insulin resistance and type 2 diabetes. *Cell* **160**, 745–758 (2015).
19. Pagnon, J. et al. Identification and functional characterization of protein kinase A phosphorylation sites in the major lipolytic protein, adipose triglyceride lipase. *Endocrinology* **153**, 4278–4289 (2012).
20. Liljenquist, J. E. et al. Effects of glucagon on lipolysis and ketogenesis in normal and diabetic men. *J. Clin. Invest.* **53**, 190–197 (1974).
21. Gravholt, C. H., Møller, N., Jensen, M. D., Christiansen, J. S. & Schmitz, O. Physiological levels of glucagon do not influence lipolysis in abdominal adipose tissue as assessed by microdialysis. *J. Clin. Endocrinol. Metab.* **86**, 2085–2089 (2001).
22. Wu, M. S. et al. Does glucagon increase plasma free fatty acid concentration in humans with normal glucose tolerance? *J. Clin. Endocrinol. Metab.* **70**, 410–416 (1990).
23. Jensen, M. D., Heiling, V. J. & Miles, J. M. Effects of glucagon on free fatty acid metabolism in humans. *J. Clin. Endocrinol. Metab.* **72**, 308–315 (1991).
24. Nichols, B. J. & Denton, R. M. Towards the molecular basis for the regulation of mitochondrial dehydrogenases by calcium ions. *Mol. Cell. Biochem.* **149–150**, 203–212 (1995).
25. Foster, D. W. Malonyl-CoA: the regulator of fatty acid synthesis and oxidation. *J. Clin. Invest.* **122**, 1958–1959 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Sample size was determined based on a power calculation to detect moderate to large differences between groups²⁶. The experiments were not formally randomized, but littermates (wild-type and knockout) were used in parallel for each experiment. Investigators were not blinded to allocation during experiments but were blinded during outcome assessment.

Animals

All protocols were approved by the Yale University Animal Care and Use Committee. Liver-specific *Insp3r1*-knockout mice were generated as previously described¹⁶, and in all experiments littermates were studied at 10–12 weeks of age. Mice were fed standard chow (Teklad no. 2018) and housed on a 12-h light/dark cycle in the Yale Animal Resources Center. To knock down ATGL in a liver-specific manner, an adeno-associated virus targeting ATGL (Vector BioLabs, 10¹² genome copies per mouse) was administered by retro-orbital injection four weeks before studies. Male mice were used for in vivo studies and both sexes were used for in vitro studies. One week before in vivo studies, mice underwent surgery under isoflurane anaesthesia to place a catheter in the jugular vein. Postsurgical recovery was demonstrated by confirming that mice had regained their presurgical body weight before any in vivo studies. Mice were fasted overnight before in vivo studies, unless otherwise stated. In the acute glucagon-infusion studies, mice were given an intravenous infusion of glucagon (5 ng per kg body weight per minute (kg⁻¹ min⁻¹)) for two hours, with tissue and plasma samples obtained after two hours of infusion. Mice were killed using intravenous pentobarbital at the conclusion of the terminal study. In the somatostatin infusion study, a 90-min jugular venous infusion of somatostatin (4 µg kg⁻¹ min⁻¹, 1:1 mix of somatostatin 14 and somatostatin 28) concurrently with [3-¹³C]lactate and [2-³H]₂glucose, as described in 'In vivo studies', was administered. During the somatostatin infusion, mice were also infused with insulin (2 mU kg⁻¹ min⁻¹) and glucagon (6 ng kg⁻¹ min⁻¹). In the acute high-dose glucagon experiment, mice were injected with 600 µg kg⁻¹ glucagon and euthanized 20 min later, after isoflurane anaesthesia. Blood was collected by cardiac puncture, and the liver was excised and freeze-clamped.

Mice infused chronically with glucagon were fed a high-fat diet (Research Diets D12492) for 4 weeks, after which they were implanted with an Alzet pump providing glucagon continuously (0.15 ng min⁻¹) for another 3.5 weeks, during which time they were continued on a high-fat diet. Mice underwent CLAMS metabolic cage analysis to assess energetics, as well as food and water intake, during the second week of glucagon infusion. After an overnight fast, three weeks after implantation of the glucagon pumps, mice underwent an intraperitoneal glucose tolerance test and were subsequently re-fed. Forty-eight hours later, after a 6-h fast, mice were euthanized under isoflurane anaesthesia for measurement of hepatic lipid content as described in 'Biochemical analysis'.

For the chronic glucagon-infusion studies in rats, 300 g male Sprague–Dawley rats were obtained from Charles River Laboratories and fed a safflower-oil-based high-fat diet (60% calories from fat) (Dyets no. 112245) for 4 weeks. During week 3 of the diet, rats underwent surgery under isoflurane anaesthesia to place catheters in the jugular vein and carotid artery, and recovery was confirmed by regaining the presurgical body weight before in vivo studies. After four weeks on a high-fat diet, rats were placed in a soft plastic harness to protect their catheters and infused continuously for 10 days with glucagon (5 ng kg⁻¹ min⁻¹, to a total volume 5 ml kg⁻¹ day⁻¹). The glucagon infusion was either continued throughout the terminal study (PINTA) or discontinued two hours before the start of the terminal study (glucose tolerance test with hepatic lipid, acetyl-CoA or glycogen measurements), in separate groups of rats, as specified in the figure legends. Rats were fasted for 8 h before being euthanized with an intravenous injection of pentobarbital.

In vivo studies

In all in vivo mouse studies, blood was collected from the tail vein, with the exception of portal-vein glucagon measurements, for which a needle was inserted into the portal vein of anaesthetized mice to collect blood. In the rat studies, blood was collected from the jugular venous catheter. In both species, samples were immediately centrifuged (12,000 rpm) to separate plasma from red blood cells. We used PINTA to measure hepatic mitochondrial fluxes in both rats and mice²⁷. In brief, rodents were infused with a 2-h primed (5 min, 3×)-continuous infusion of [3-¹³C]lactate (40 µmol kg⁻¹ min⁻¹) and either [1,2,3,4,5,6,6-²H]₂glucose (0.1 mg kg⁻¹ min⁻¹) or [3-³H]glucose (0.1 µCi kg⁻¹ min⁻¹). At the conclusion of the study, animals were euthanized with an intravenous injection of pentobarbital.

In mice, hepatic glycogenolysis was assumed to be negligible owing to their prolonged (16 h) fasted state and their low hepatic glycogen content (Extended Data Figs. 1f, 5i, 7p). HGP was measured by determining the specific activity of [3-³H]glucose in plasma using a scintillation counter, and the V_{PC}/V_{HGP} ratio was calculated using the equation

$$\frac{V_{PC}}{V_{HGP}} = \frac{G2}{XFE^2}$$

in which $G2$ represents the $[m+2]$ glucose enrichment corrected for any $[m+2]$ glucose synthesized from ¹³C₂-labelled trioses: corrected $[m+2]$ glucose = $G2$ = measured $[m+2]$ glucose – $2 \times C4C5C6[m+2]$ glucose, and XFE represents the fractional triose enrichment:

$$XFE = \frac{1}{1 + \frac{G1}{2 \times G2}}$$

in which $G1$ represents the measured $[m+1]$ glucose, and $G2$ is as described above. To calculate absolute V_{PC} , we multiplied the measured HGP by the ratio V_{PC}/V_{HGP} . The ratio of hepatic V_{PC}/V_{CS} was calculated as

$$\frac{V_{PC}}{V_{CS}} = \frac{[5-^{13}\text{C}]\text{glucose}}{2 \times [4-^{13}\text{C}]\text{glucose}} - 1$$

and the absolute V_{CS} was calculated by dividing V_{PC} by V_{PC}/V_{CS} . The derivations of each equation are described in detail in a previous study²⁷. We corrected for the possible contribution of [¹³C]bicarbonate to label the TCA cycle, as previously described²⁸.

We measured the ratio

$$\frac{V_{PDH}}{V_{CS}} = \frac{[4-^{13}\text{C}]\text{glutamate}}{[3-^{13}\text{C}]\text{alanine}}$$

and calculated the absolute V_{PDH} by multiplying this ratio by the measured V_{CS} . Finally, the ratio of V_{PK} —assuming minimal malic enzyme flux—to ($V_{PC} + V_{PDH}$) was calculated as (ref.²⁹)

$$\frac{V_{PK}}{V_{PC} + V_{PDH}} = \frac{[2-^{13}\text{C}]\text{alanine}}{[5-^{13}\text{C}]\text{glucose}}$$

Absolute V_{PK} rates were then determined by multiplying $V_{PK}/(V_{PC} + V_{PDH})$ by the sum of V_{PC} and V_{PDH} . As has previously been described²⁷, V_{PC}/V_{CS} can be expanded to account for pyruvate recycling (ref.²⁹):

$$\frac{V_{PC} + \frac{1}{2}V_{PK}}{V_{CS}} = \frac{[5-^{13}\text{C}]\text{glucose}}{2 \times [4-^{13}\text{C}]\text{glucose}} - 1$$

In this study, we measured a maximum $V_{PK}/(V_{PC} + V_{PDH})$ of 0.4, indicating that the maximal V_{PK}/V_{PC} is 0.4. A V_{PK}/V_{PC} at this maximal value would generate a 17% underestimation of V_{PC}/V_{CS} . Ex vivo NMR analysis was

used to confirm PINTA measurements of flux ratios as has previously been described²⁷.

In rats, HGP was measured by measuring the plasma glucose [$m + 7$] atom per cent enrichment (APE) by gas chromatography–mass spectrometry (GC–MS) and using these data to calculate HGP according to the equation $HGP = (\text{tracer APE/plasma APE} - 1) \times \text{infusion rate}$. All other flux ratios and absolute fluxes were measured using the equations given above.

In the glucose tolerance tests, rodents were injected with 1 g kg^{-1} 50% dextrose (rats) or 10% dextrose (mice) intraperitoneally. Blood samples were taken through the venous catheter (rats) or by tail bleeding (mice) for measurement of plasma glucose and insulin concentrations as described in ‘Biochemical analysis’.

Biochemical analysis

Glucose concentrations in plasma and in cell-culture media were measured using the YSI Glucose Analyzer. Plasma insulin was measured by enzyme-linked immunosorbent assay (ELISA) (Mercodia), and glucagon was measured in samples immediately spiked with aprotinin ($0.5 \text{ mg } \mu\text{l}^{-1}$) whole blood by radioimmunoassay by the Yale Diabetes Research Core. Plasma NEFA concentrations were measured using the Wako NEFA assay, and plasma glycerol³⁰ and plasma and liver amino acid concentrations¹⁷ were measured by GC–MS. Liver acetyl- and malonyl-CoA¹⁸ (standard curve $R^2 = 0.999$ and 0.999), long-chain acyl-CoA³¹, DAG³¹ and ceramide³¹ concentrations were measured by liquid chromatography with tandem mass spectrometry, hepatic glycogen content was measured following amyloglucosidase digestion³² and TAG concentrations were measured enzymatically³³. cAMP concentrations were measured using the Enzo Life Sciences Direct cAMP ELISA. Protein concentrations were measured by western blot, using antibodies from Santa Cruz (PC, PEPCCK and CAMKII), Cell Signaling (ATGL, pCAMKIV, CAMKIV, pCRTC2, CRTC2, pACC, ACC, pAMPK, AMPK, pHSL, HSL, GAPDH and β -actin), Novus Biologicals (pCAMKII), Abcam (pATGL), LSBio (pINSP3R1) and BD Transduction Laboratories (INSP3R3 and PKC ϵ). The antibody to pATGL was kindly provided by H. S. Sul, and the antibody to total INSP3R1 was custom-made using an epitope against the last 18 amino acids of INSP3R1. The antibody to INSP3R2 was generously donated by the Wojcikiewicz Laboratory. The antibody to CRTC2 as well as hepatocyte control samples (wild-type \pm glucagon and *Crtc2* knockout) for antibody validation were a kind gift of M. Montminy. To measure protein expression in isolated hepatocytes, a Mitochondria Isolation Kit (Thermo Scientific) was used to separate cells into cytoplasmic and crude mitochondrial fractions, the latter containing both mitochondria and closely apposed mitochondria-associated proteins. The final spin was conducted at 1,000 rpm for 15 min at 4°C to obtain a cleaner mitochondrial fraction with reduced lysosomal contamination. Mitochondria were lysed in 2% CHAPS in tris-buffered saline (TBS), and the protein concentration was measured with the bicinchoninic acid assay (Thermo Scientific). Samples for all western blot analyses were loaded on 4 to 12% gradient gels (NuPage gels, Life Technologies) and run with MOPS buffer. Protein was transferred to polyvinylidene difluoride membranes by wet transfer. After blocking with 5% milk in TBS with 0.1% Tween 20, primary antibodies were applied overnight. Gluconeogenic gene mRNA expression was measured by quantitative PCR³⁴ using primers with the following sequences: *Actb*: forward (F), CCAGATCATGTTTGAGACCTTC, reverse (R), CATGAGGTAGTCTGTCAGGTCC; *Pepck* (also known as *Pck1*): F, CAGGAAGTGAGGAAGTTGTGG, R, ATGACACCCTCCTCCTGCAT; *G6Pase* (also known as *G6pc*): F, GAAGGCCAAGATGGTGTGA, R, TGCAGCTCTTGCGGTACATG; and *Pc* (also known as *Pcx*): F, AGATGCACTTCCATCCCAAG, R, CCTTGGTCACGTGAACCTTT

In vitro studies

Primary hepatocytes were isolated by the Yale Liver Center from wild-type and *Insp3r1*-knockout mice, and plated on glass coverslips coated

with rat collagen type I. Cells were loaded with the cytosolic Ca^{2+} indicator dye Fluo-4 AM (Thermo Fisher Scientific), or the mitochondrial-matrix-targeted Ca^{2+} indicator Rhod-2 AM (Thermo Fisher Scientific), for 30 min at 37°C . Ca^{2+} imaging experiments were performed in 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES)-buffered balanced salt solution (25 mM HEPES, 121 mM NaCl, 4.7 mM KCl, 1.2 mM MgSO_4 , 1.2 mM KPO_4 , 5 mM NaHCO_3 , 2.0 mM CaCl_2 , 10 mM glucose, pH 7.4) with or without the PKA inhibitor H-89 (25 μM) or the PLC inhibitor U73122 (10 μM) 2–4 h following the initial plating. Coverslips were transferred to a custom-built perfusion chamber on the stage of a Zeiss LSM 710 confocal microscope (Carl Zeiss Microscopy). Cytosolic and mitochondrial signals were monitored in Fluo-4 AM and Rhod-2-AM-loaded cells, respectively, using stimulation with 100-nM glucagon (Sigma-Aldrich) or 100-nM vasopressin (Sigma-Aldrich) with a $20\times$ objective lens. Changes in fluorescence were normalized to the initial fluorescence before addition of agonist and were expressed as F/F_0 .

For in vitro glucose production and lipolysis studies, primary mouse hepatocytes were isolated by the Yale Liver Center. Following removal of cell debris by Percoll density gradient centrifugation, cells were plated on 6-well collagen-I-coated dishes (4.0×10^5 cells per well) in 2 ml recovery medium (DMEM high glucose containing 10% FBS, 2% penicillin–streptomycin, 100 nM dexamethasone, 1 nM insulin and 10 mM HEPES). After incubation for 6 h at 37°C and 5% CO_2 , the attached cells were washed once in $1\times$ PBS and then incubated overnight in 2 ml low-glucose culture medium (DMEM low glucose supplemented with 10% FBS, 2% penicillin–streptomycin and 10 mM HEPES) for glucose production studies, or serum-free low-glucose culture medium (DMEM low glucose supplemented with 0.5% fatty acid free BSA, 2% penicillin–streptomycin and 10 mM HEPES) for lipolysis assays. The next morning, cells were washed twice in $1\times$ PBS and culture media replaced with 2 ml glucose production medium (DMEM base medium supplemented with 0.5% fatty acid free BSA, 20 mM sodium lactate, 2 mM sodium pyruvate and 10 mM HEPES, pH 7.4) or serum-free low glucose culture medium (lipolysis assay) in the presence of 100 nM glucagon or vehicle control. After incubation for 8 h at 37°C and 5% CO_2 , cell culture medium was collected for analysis of glucose, NEFA and glycerol concentrations as described in ‘Biochemical analysis’. V_{PC} was determined by measuring $V_{\text{PC}}/V_{\text{EGP}}$ by PINTA as described in ‘In vivo studies’ using 300 μl of the collected medium and multiplying this ratio by the measured glucose production rate. In a subset of studies, cells were incubated for 8 h during the glucose production assay in medium containing one of the following agents (all dissolved in 0.5% DMSO), or 0.5% DMSO vehicle: 20 μM ET-18-OCH₃ (Santa Cruz), 100 μM U-73122 (Sigma), 30 nM thapsigargin (Sigma), 100 nM vasopressin (Sigma), 50 μM 2-APB (R&D Systems), 70 μM caffeine (Sigma) or 1 mM malic enzyme inhibitor hydroxymalonate (Sigma). To inhibit PKA, hepatocytes were incubated in H-89 dichloroacetate hydrate (Sigma, 25 μM) dissolved in medium, or medium lacking H-89 as a control. In the insulin treatment studies, hepatocytes were incubated in 1 nM insulin with 5% BSA and glucose production medium for the duration of the glucose production assay. In the in vitro atglistatin study, hepatocytes were incubated in 10 μM atglistatin in 0.1% EtOH, or 0.1% EtOH vehicle. All values were normalized to total protein content determined from whole-cell lysates by a BCA protein assay (Thermo Fisher Scientific) according to the manufacturer’s instructions and expressed as fold change versus vehicle-treated cells.

To assess oxygen consumption, primary mouse hepatocytes were isolated by the Yale Liver Center and plated on collagen-I-coated XF24 V7 cell culture plates (1.2×10^4 cells per well) in 2 ml recovery medium, as previously described³⁵. Following incubation for 6 h at 37°C and 5% CO_2 , the cells were washed twice with 1 ml low-glucose culture medium and incubated overnight in 250 μl low-glucose culture medium at 37°C and 5% CO_2 . The following morning, cells were washed with XF24 assay medium (DMEM base medium containing 1.0 mM pyruvate, 2 mM glutamine and 5.5 mM glucose, pH 7.4). Five hundred microlitres of XF24

Article

assay medium was added to each well and plates were equilibrated at 37 °C for 1 h. Four measurements of basal oxygen consumption rates (picomoles per minute) were recorded on a Seahorse Bioscience XFe 24 Analyzer (Seahorse Biosciences) using an instrument protocol of 3-min mix, 2-min wait and 3-min measure. After baseline measurements, glucagon (or vehicle) was injected at a final concentration of 100 nM and oxygen consumption was recorded using the same instrument protocol. Ten measurements were taken following injection and the average of eight measurements was used for subsequent analyses. Oxygen consumption rates were normalized to total protein content and expressed as fold change compared to vehicle-treated cells.

Statistics and reproducibility

Group sizes were chosen to detect moderate-to-large (>40%) differences with approximately 40% standard deviations. Comparisons were performed using GraphPad Prism 7. Each in vivo experiment was performed with the number of replicates specified in the figure legends. All data obtained are shown, with the exception of several western blots in which case one representative image is shown; all data are shown in the quantification. The two-tailed paired (when comparing the same animals under multiple conditions) or unpaired Student's *t*-test (when comparing different animals) was used to compare datasets. *P* values less than 0.05 were considered significant. Comparisons were performed using GraphPad Prism 7.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated and analysed in this study are available in the Article and its Supplementary Information files, which include a table containing all raw data. Source Data for Figs. 1–3, Extended Data Figs. 1–8 are included with the paper.

26. Kadam, P. & Bhalerao, S. Sample size calculation. *Int. J. Ayurveda Res.* **1**, 55–57 (2010).
27. Perry, R. J. et al. Non-invasive assessment of hepatic mitochondrial metabolism by positional isotopomer NMR tracer analysis (PINTA). *Nat. Commun.* **8**, 798 (2017).
28. Petersen, K. F., Dufour, S., Cline, G. W. & Shulman, G. I. Regulation of hepatic mitochondrial oxidation by glucose–alanine cycling during starvation in humans. *J. Clin. Invest.* **129**, 4671–4675 (2019).
29. Perry, R. J. et al. Propionate increases hepatic pyruvate cycling and anaplerosis and alters mitochondrial metabolism. *J. Biol. Chem.* **291**, 12161–12170 (2016).
30. Perry, R. J. et al. Mechanism for leptin's acute insulin-independent effect to reverse diabetic ketoacidosis. *J. Clin. Invest.* **127**, 657–669 (2017).
31. Yu, C. et al. Mechanism by which fatty acids inhibit insulin activation of insulin receptor substrate-1 (IRS-1)-associated phosphatidylinositol 3-kinase activity in muscle. *J. Biol. Chem.* **277**, 50230–50236 (2002).
32. Jurczak, M. J. et al. SGLT2 deletion improves glucose homeostasis and preserves pancreatic β -cell function. *Diabetes* **60**, 890–898 (2011).
33. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
34. Camporez, J. P. et al. ApoA5 knockdown improves whole-body insulin sensitivity in high-fat-fed mice by reducing ectopic lipid content. *J. Lipid Res.* **56**, 526–536 (2015).
35. Camporez, J. P. et al. Cellular mechanisms by which FGF21 improves insulin sensitivity in male mice. *Endocrinology* **154**, 3099–3109 (2013).

Acknowledgements We thank M. Montminy, T. Sonntag and Y.-S. Yoon for kindly providing the CRTC2 antibody and control samples, and for their advice on the interpretation of the CRTC2 blots; H. S. Sui for generously providing the pATGL antibody; D. Vatner for his advice on these studies; and X. Ma, J. Dong, W. Zhu, M. Kahn, K. Harry and M. Batsu for their expert technical assistance. These studies were funded by grants from the United States Public Health Service (R01 DK113984, P30 DK059635, P30 DK034989, T32 DK101019, K99/R00 CA215315, R01 NS087568, UL1TR000142, T32 DK007058 and F32 DK114954). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions The study was designed by R.J.P. and G.I.S. Data were collected and analysed by R.J.P., D.Z., M.T.G., A.L.B., L.G., A.R.N., A.R.-C., Y.W., L.P., S.D., Y.Z., X.-M.Z., G.M.B., K.T., Y.N., K.F.P., G.W.C., B.E.E. and M.H.N. The manuscript was written by R.J.P. and G.I.S. with contributions and approval from all authors.

Competing interests The authors declare no competing interests.

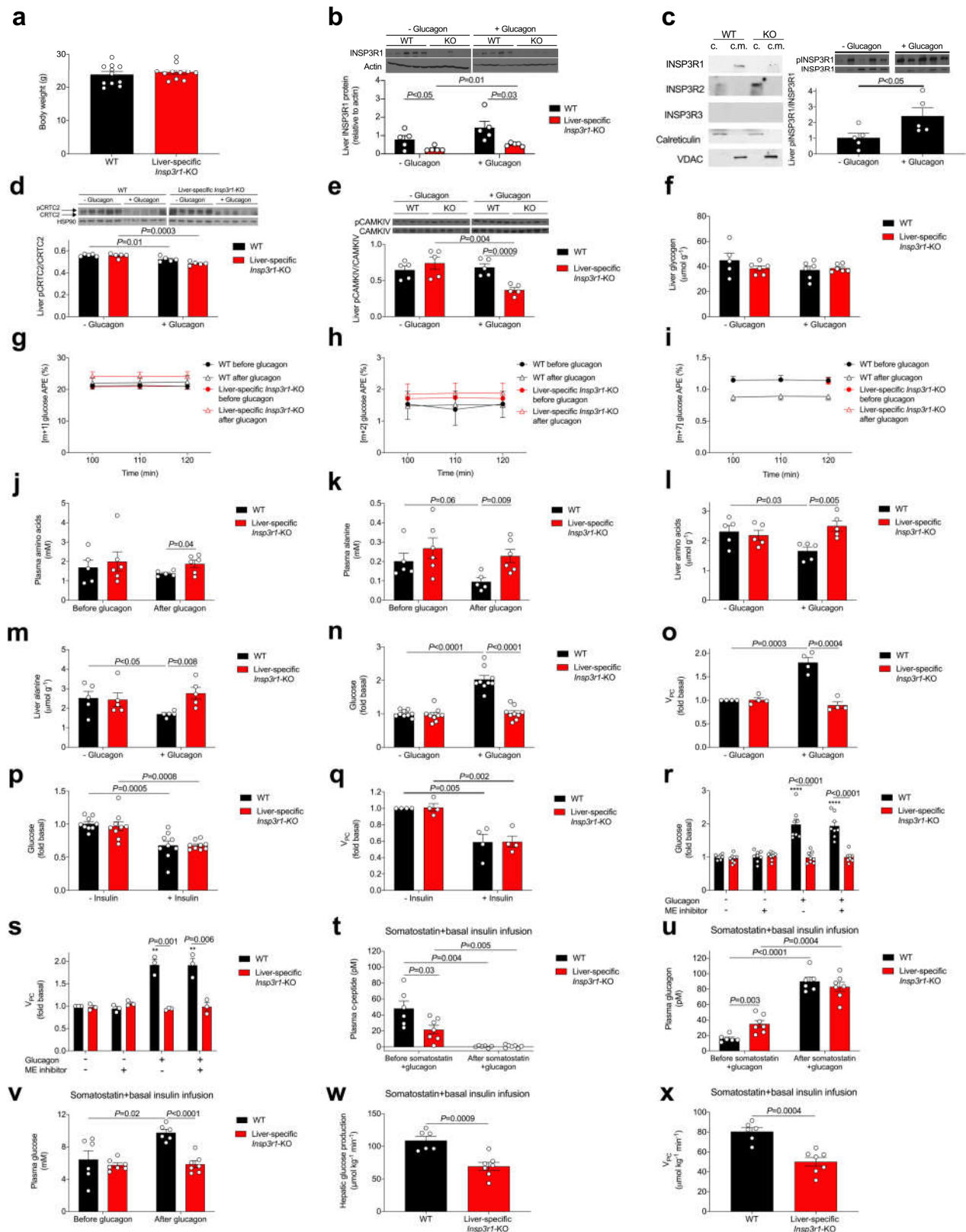
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2074-6>.

Correspondence and requests for materials should be addressed to G.I.S.

Peer review information Nature thanks Ilya Bezprozvanny and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

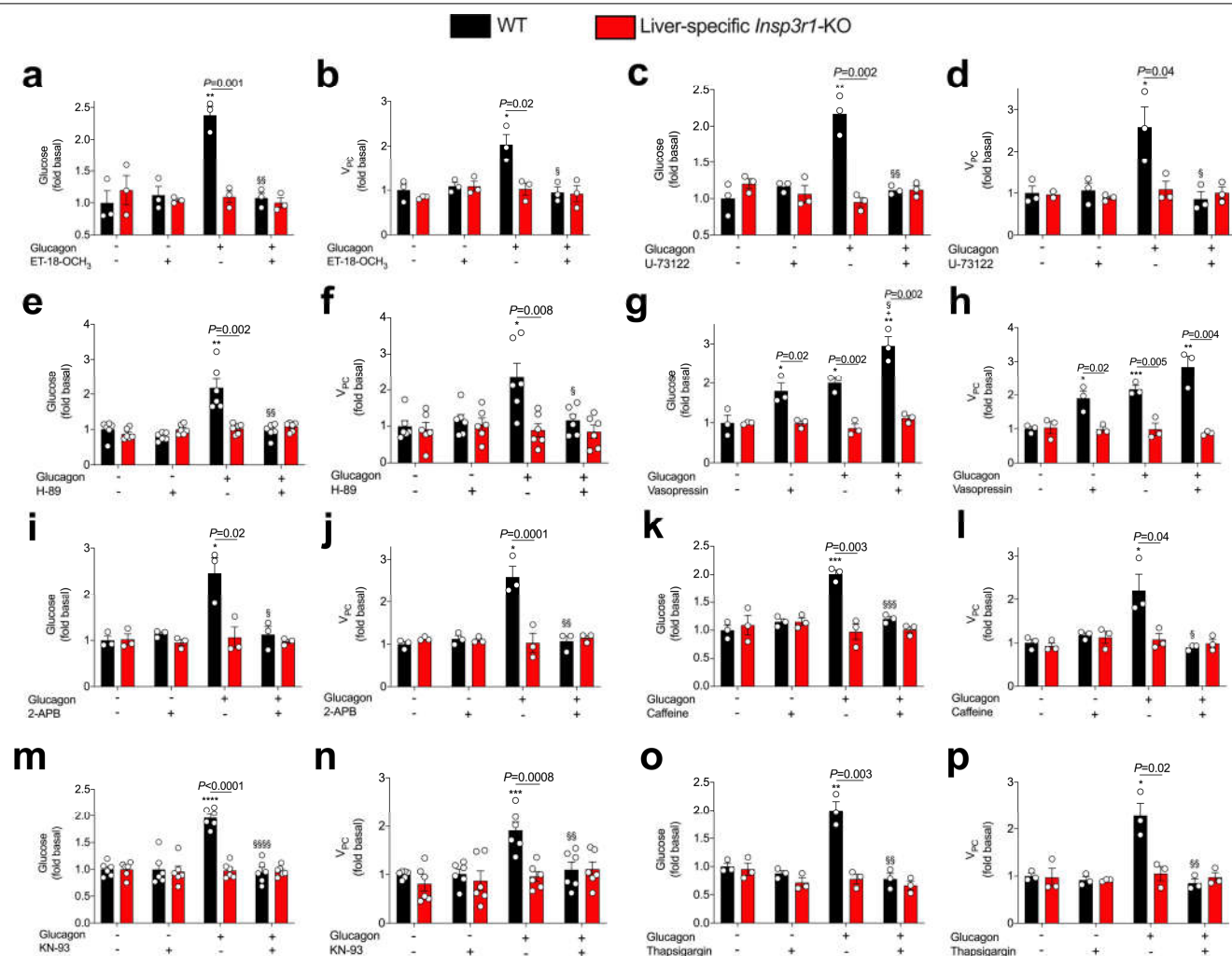


Extended Data Fig. 1 | See next page for caption.

Article

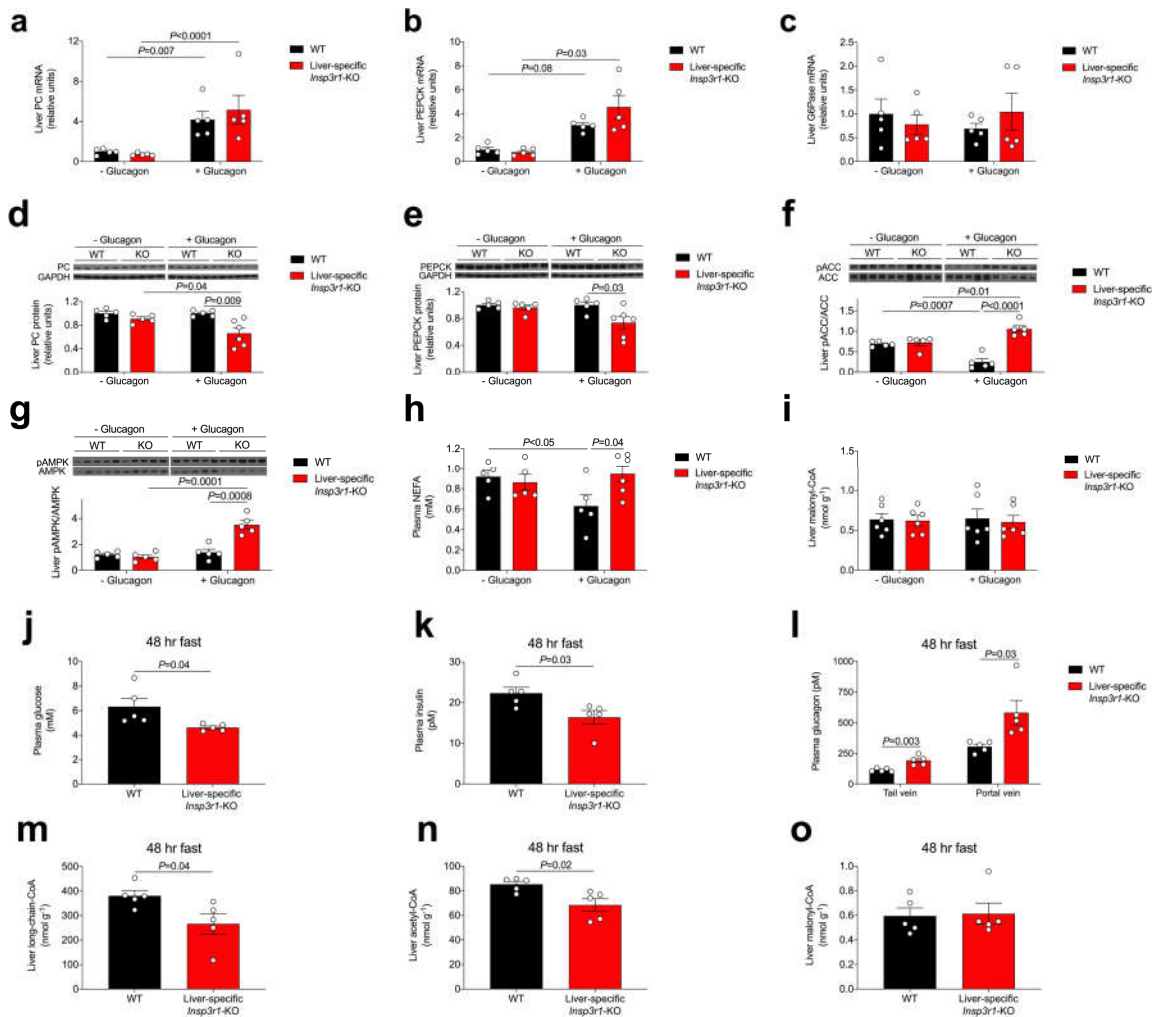
Extended Data Fig. 1 | Glucagon acutely stimulates hepatic gluconeogenesis by increasing hepatic acetyl-CoA content and PC flux. **a**, Body weight ($n=11$). **b**, Expression of INSP3R1 protein in the liver ($n=5$). Blots in **b**, **c**, **e** and Figs. 1f, 2a, Extended Data Figs. 3f, g, 4a were stripped and reprobed for all proteins of interest. $^{*}P<0.05$ versus *Insp3r1*-knockout mice not treated with glucagon. **c**, INSP3R protein expression in cytosolic (c.) and crude mitochondrial (c.m.) fractions from primary hepatocytes, in which VDAC was examined as a marker for mitochondrial protein content, and calreticulin as a marker for non-mitochondria-associated membrane protein. Right, phosphorylation of INSP3R1 in the liver of mice infused with glucagon ($n=5$). **d**, Phosphorylation of CRT2 in the liver ($n=5$). The CRT2 phosphorylation gel was stripped and reprobed to assess HSP90 (loading control). **e**, Phosphorylation of CAMKIV in the liver, with or without a 2-h acute infusion of glucagon ($n=5$). **f**, Liver glycogen content ($n=5$ wild type – glucagon, otherwise $n=6$). No differences were observed using one-way analysis of variance with Bonferroni's multiple comparisons test. **g–i**, Plasma [$m+1$], [$m+2$] and [$m+7$] glucose enrichment during a 2-h infusion of [$3-^{13}C$] lactate and [$2H_2$] glucose ($n=5$ wild type and 6 knockout, with the exception of **i**, for which $n=4$ wild type + glucagon at 100 and 110 min). **j**, **k**, Plasma total amino acid and alanine concentrations ($n=5$ wild type and 6 knockout). In **j**, **k**, groups were compared before and after

glucagon treatment by two-tailed paired Student's *t*-test, and genotypes were compared by two-tailed unpaired Student's *t*-test. **l**, **m**, Liver total amino acid and alanine concentrations ($n=5$). **n**, **o**, In vitro glucose production ($n=9$) and V_{PC} ($n=4$) in isolated hepatocytes. **p**, **q**, In vitro glucose production ($n=9$) and V_{PC} ($n=4$) in isolated hepatocytes with and without 150 pM insulin. Basal data (no insulin) are duplicated from **n** and **o**. **r**, **s**, In vitro glucose production ($n=8$) and V_{PC} ($n=3$) in isolated hepatocytes with and without a malic enzyme (ME) inhibitor. $^{**}P<0.01$, $^{****}P<0.0001$ versus wild type – glucagon – ME inhibitor. **t–v**, Plasma connecting peptide, glucagon and glucose concentrations in mice ($n=6$ wild type and 7 knockout) treated with somatostatin, basal insulin and glucagon. Comparisons before and after glucagon treatment used a two-tailed paired Student's *t*-test. **w**, **x**, Endogenous glucose production and V_{PC} ($n=6$ wild type and 7 knockout). In all panels, comparisons with and without glucagon, insulin or malic enzyme inhibitor, and wild type versus knockout comparisons were performed using a two-tailed unpaired Student's *t*-test, unless otherwise stated. In all panels in which comparisons were performed (that is, all panels with the exception of **g**, **i**), if no *P* value is shown, groups were not significantly different. In all panels, mean \pm s.e.m. is shown. All *n* values refer to numbers of mice.



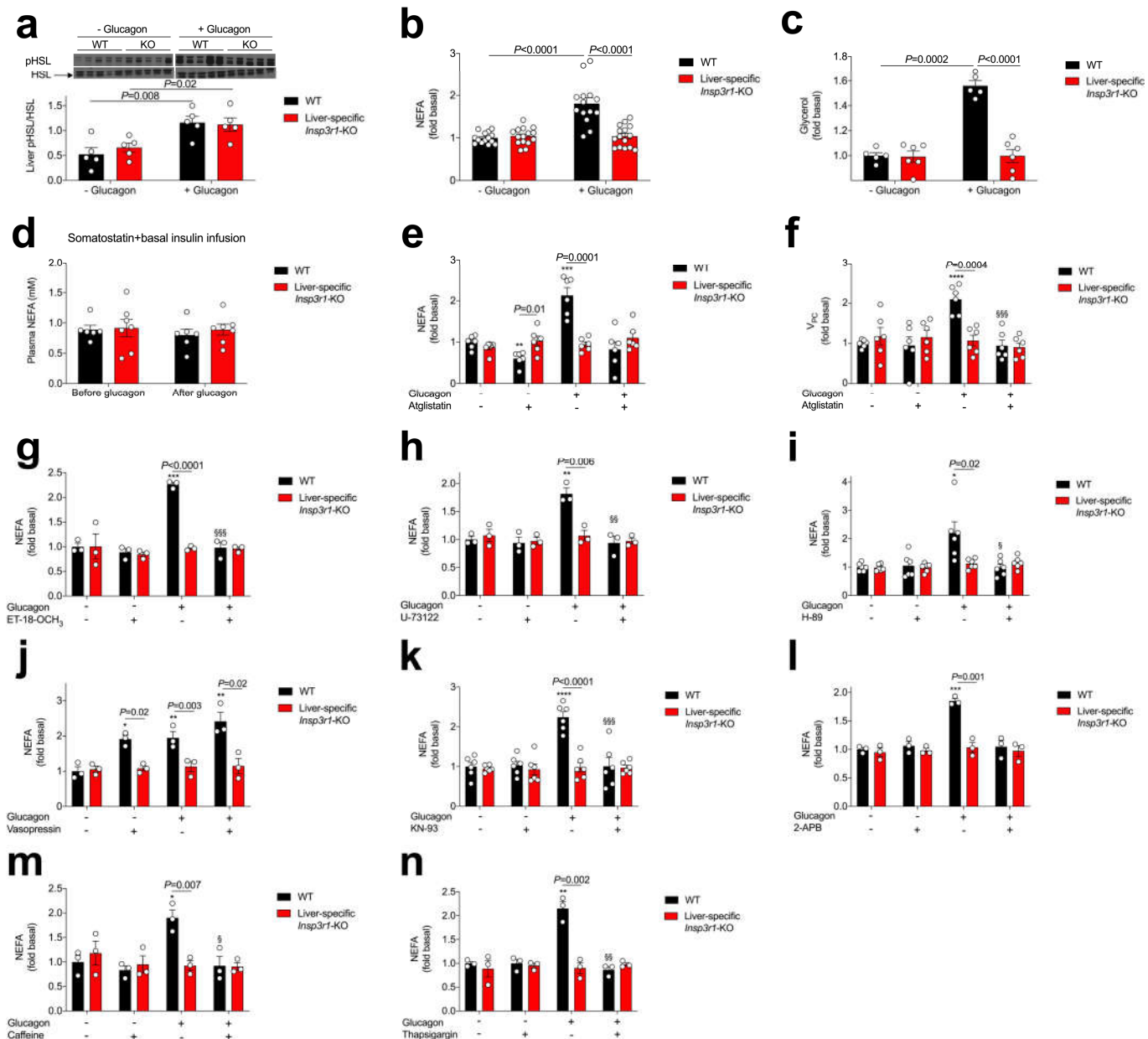
Extended Data Fig. 2 | Glucagon-stimulated glucose production requires activation of the PLC and PKA pathways, converging to activate INSP3 signalling. **a–p**, In vitro glucose production and V_{pc} in isolated hepatocytes with and without ET-18-OCH₃ ($n=3$) (**a**, **b**), U-73122 ($n=3$, except for knockout – glucagon-U-73122 in **d** ($n=2$)) (**c**, **d**), H-89 ($n=6$) (**e**, **f**), vasopressin ($n=3$) (**g**, **h**), 2-APB ($n=3$) (**i**, **j**), caffeine ($n=3$) (**k**, **l**), KN-93 ($n=6$) (**m**, **n**) and

thapsigargin ($n=3$) (**o**, **p**). In all panels, $*P<0.05$, $**P<0.01$ and $***P<0.001$ versus wild type – glucagon – drug; $\$P<0.05$, $\$\$P<0.01$ and $\$\$\$P<0.001$ versus wild type + glucagon – drug by two-tailed unpaired Student's t -test. If no statistical comparison is denoted, the groups were not significantly different. Mean \pm s.e.m. is shown.



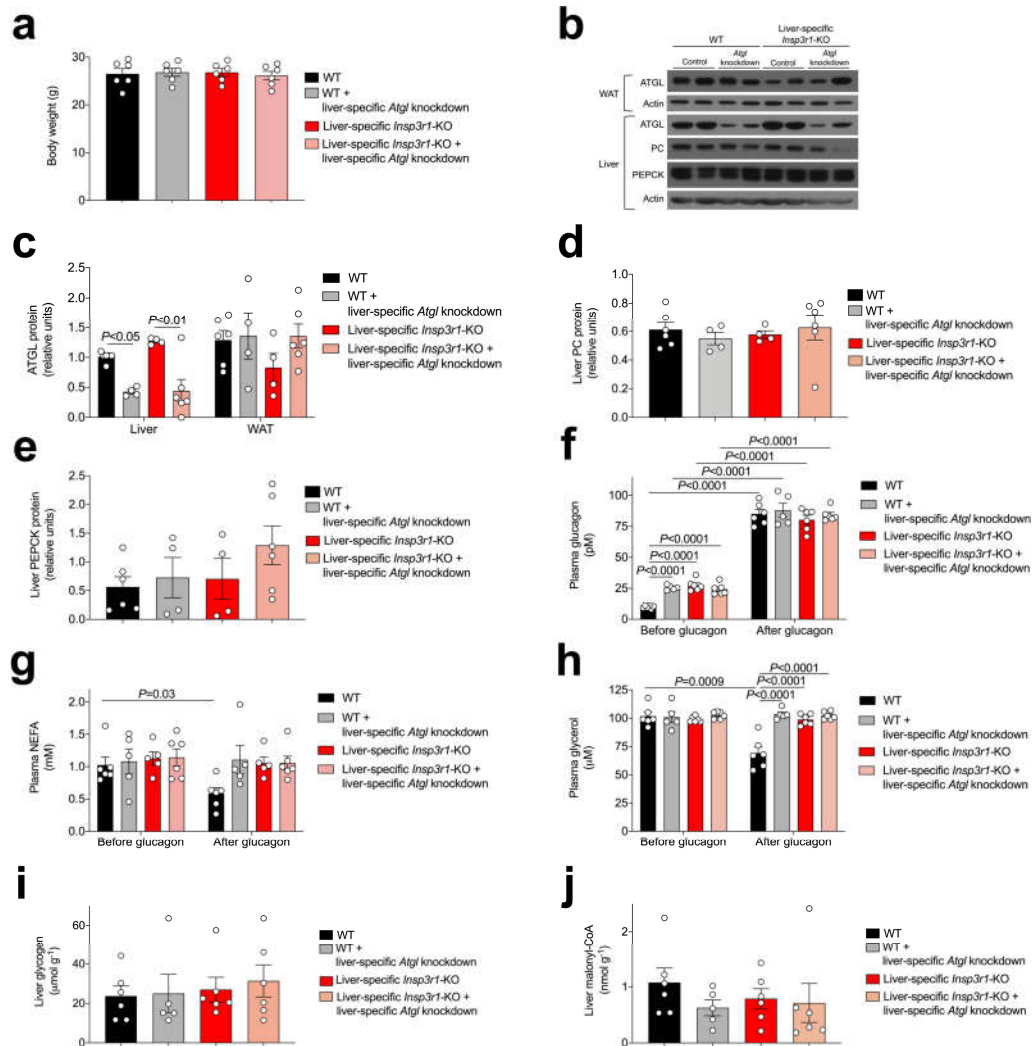
Extended Data Fig. 3 | Glucagon stimulates HGP independently of transcriptional regulation, and has a key role in the maintenance of blood glucose levels during a prolonged fast. **a–c**, Liver *Pc*, *Pepck* and *G6Pase* mRNA expression ($n=5$). **d, e**, Liver *Pc* and *PEPCK* protein ($n=5$, except for knockout + glucagon ($n=6$)). **f, g**, Liver pACC/ACC and pAMPK/AMPK ratios ($n=5$). Blots in **f, g** and Figs. 1f, 2a, Extended Data Figs. 1b, c, e, 4a were stripped and reprobed for all proteins of interest. **h**, Plasma NEFA ($n=5$, except for

knockout + glucagon ($n=6$)). **i**, Liver malonyl-CoA ($n=6$). **j–l**, Plasma glucose, insulin and glucagon concentrations in mice fasted for 48 h ($n=5$). **m–o**, Liver long-chain-, acetyl- and malonyl-CoA content ($n=5$). In all panels, genotypes and groups \pm glucagon were compared using a two-tailed unpaired Student's *t*-test. If no statistical comparison is denoted, the groups were not significantly different. In all panels, mean \pm s.e.m. is shown. All *n* values refer to numbers of mice.



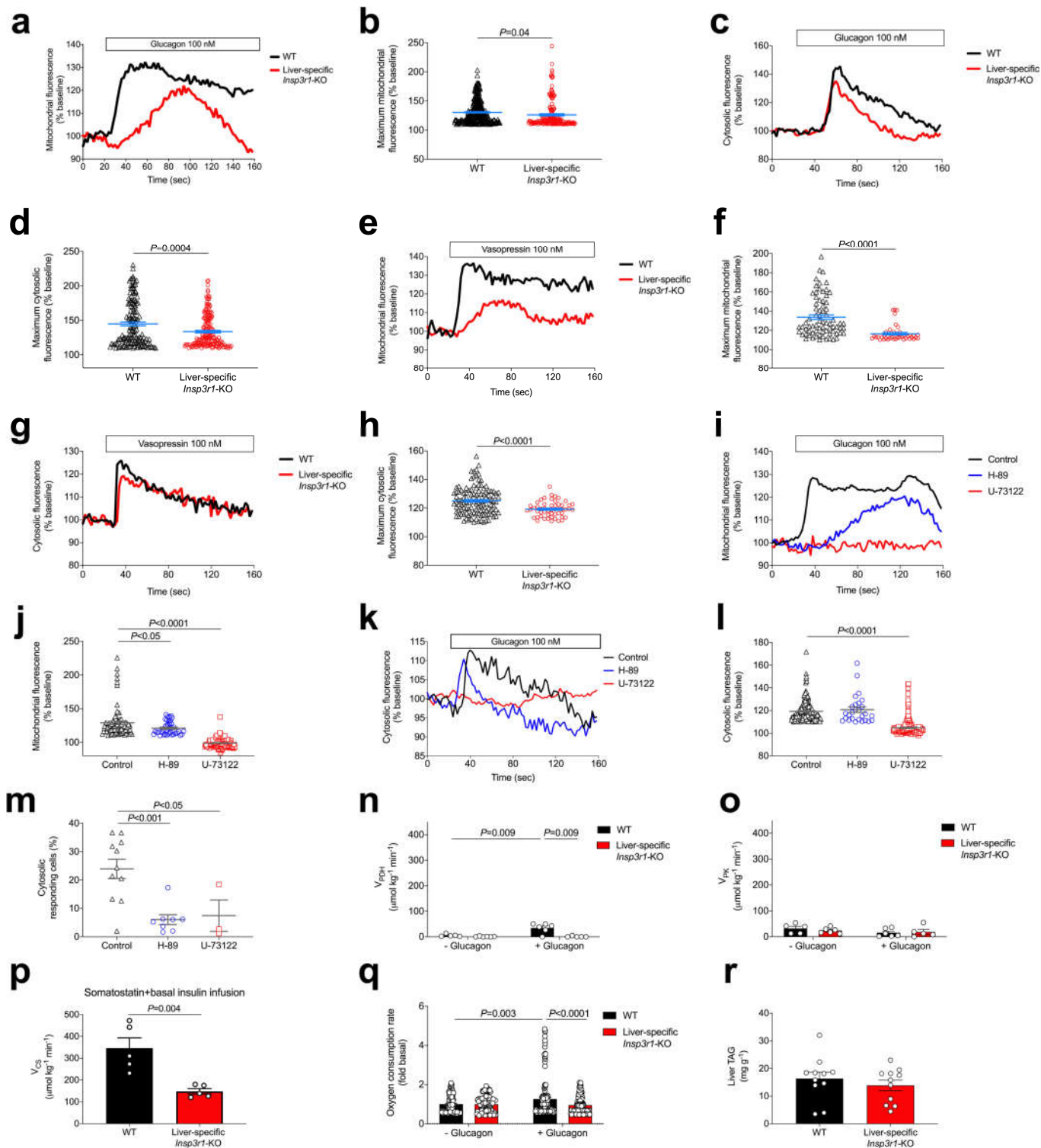
Extended Data Fig. 4 | Glucagon acutely stimulates gluconeogenesis by activating intrahepatic, but not white adipose tissue, lipolysis. **a**, Liver HSL phosphorylation ($n=5$). These blots, and blots in Figs. 1f, 2a, Extended Data Figs. 1b, c, e, 3f, g, were stripped and reprobed for all proteins of interest. In **a–c**, groups were compared using a two-tailed unpaired Student's *t*-test. **b, c**, In vitro NEFA ($n=14$ wild type and 15 knockout) and glycerol production ($n=5$ wild type and 6 knockout) from isolated hepatocytes. **d**, Plasma NEFA concentrations in mice treated with somatostatin, replacement basal insulin and glucagon ($n=6$ wild type and 7 knockout). No significant differences were observed between genotypes using a two-tailed unpaired Student *t*-test, or before versus

after glucagon treatment using a two-tailed paired Student's *t*-test. **e, f**, NEFA production and V_{FC} in isolated hepatocytes incubated in the ATGL inhibitor atglstatin ($n=6$). **g–n**, NEFA production from isolated hepatocytes treated with ET-18-OCH₃ ($n=3$), U-73122 ($n=3$), H-89 ($n=6$), vasopressin ($n=3$), KN-93 ($n=6$), 2-APB ($n=3$), caffeine ($n=3$) and thapsigargin ($n=3$). In all panels, $*P<0.05$, $**P<0.01$ and $***P<0.001$ versus the same genotype – glucagon – drug; $\$P<0.05$, $\$\$P<0.01$ and $\$\$\$P<0.001$ versus the same genotype + glucagon – drug by two-tailed unpaired Student's *t*-test. If no statistical comparison is denoted, the groups were not significantly different. Error bars represent s.e.m. All *n* values refer to numbers of mice.



Extended Data Fig. 5 | Glucagon requires INSP3-mediated intrahepatic lipolysis to promote V_{pc} and hepatic gluconeogenesis. **a**, Body weight in mice treated with an adeno-associated virus to knock down liver *Atgl* ($n = 6$). **b**, Representative western blots. Blots from the same tissue (liver or white adipose tissue (WAT)) were stripped and reprobed for all proteins shown. **c**, White adipose tissue and liver ATGL protein expression. In **c–e**, $n = 6$, except for wild type + *Atgl* knockdown and knockout ($n = 4$). **d**, **e**, Hepatic PC and PEPCK protein expression. **f–h**, Plasma glucagon ($n = 6$, except for wild type + *Atgl* knockdown ($n = 5$), NEFA ($n = 6$, except for wild type + *Atgl* knockdown and

knockout ($n = 5$), and glycerol concentrations ($n = 6$, except for wild type + *Atgl* knockdown ($n = 5$) in mice treated with an adeno-associated virus to knock down ATGL in a liver-specific manner. Groups were compared before and after glucagon treatment by two-tailed unpaired Student's *t*-test. **i**, **j**, Liver glycogen and malonyl-CoA content ($n = 6$, except for wild type + ATGL knockdown ($n = 5$)). All comparisons were performed using a two-tailed unpaired Student's *t*-test, unless otherwise stated. If no statistical comparison is denoted, the groups were not significantly different. Error bars represent s.e.m. All *n* values refer to numbers of mice.



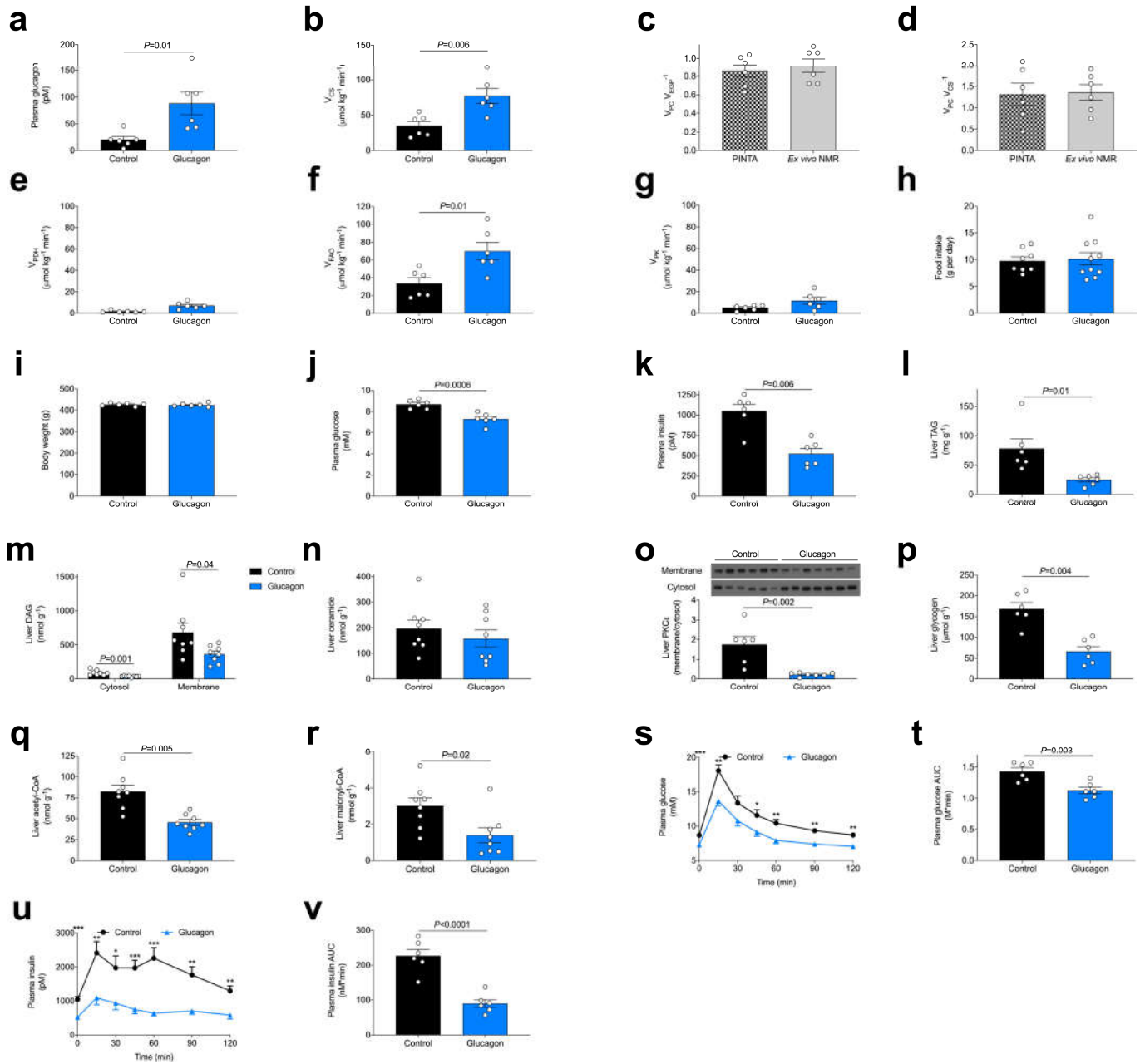
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Glucagon stimulates mitochondrial oxidation

through hepatocellular calcium signalling. **a**, Representative mitochondrial response to glucagon, which was added where denoted by the 'glucagon' bar. **b**, Maximum mitochondrial response to glucagon ($n = 324$ wild type and 167 knockout). Groups in **b**, **d**, **f** and **h** were compared by two-tailed unpaired Student's *t*-test. **c**, Representative cytosolic response to glucagon. **d**, Maximum cytosolic response to glucagon ($n = 146$ wild type and 175 knockout). **e**, Representative mitochondrial response to the INSP3R agonist vasopressin. **f**, Amplitude of the mitochondrial response to vasopressin ($n = 73$ wild type and 42 knockout). **g**, Representative cytosolic response to vasopressin. **h**, Amplitude of the cytosolic response to vasopressin ($n = 119$ wild type and 53 knockout). **i**, Representative mitochondrial responses to glucagon (added where denoted by the bar) in wild-type hepatocytes incubated in the PKA inhibitor H-89 or the PLC inhibitor U-73122. **j**, Amplitude of the mitochondrial response ($n = 79$ control, 39 H-89 and 127 U-73122). Groups were compared to the control in **j**, **l**, **m** using a two-tailed unpaired Student's *t*-test.

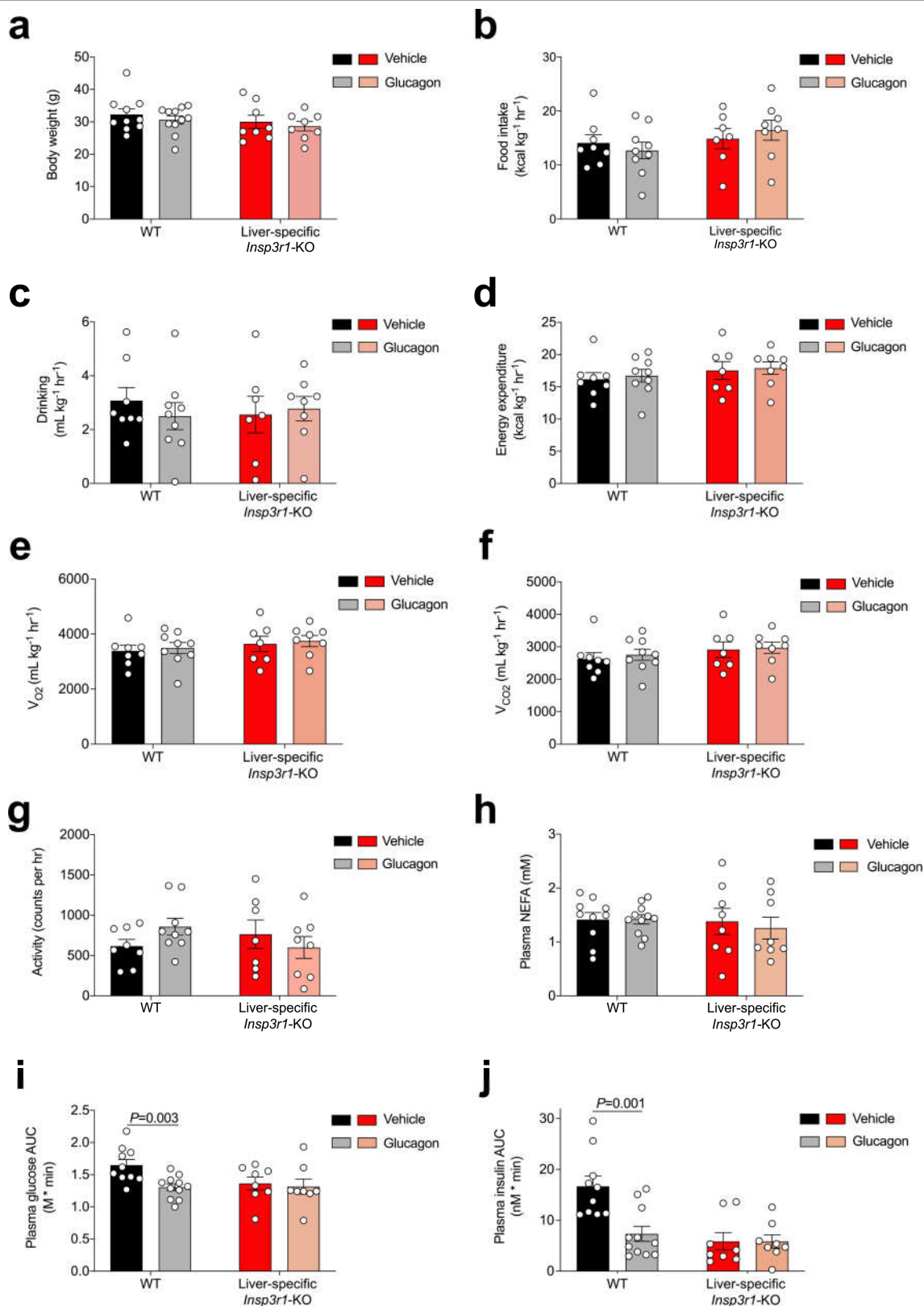
k, Representative cytosolic responses to glucagon \pm H-89 or U-73122.

l, Amplitude of the cytosolic response ($n = 182$ control, 27 H-89 and 187 U-73122). **m**, Percentage of cells with a cytosolic response ($>110\%$ baseline) to glucagon \pm H-89 or U-73122 ($n = 11$ control, 8 H-89 and 3 U-73122). **n, o**, Hepatic V_{PDH} and V_{PK} in vivo ($n = 5$ wild type – glucagon, 6 wild type + glucagon, 6 knockout – glucagon and 5 knockout + glucagon). In **n, o**, groups were compared (with versus without glucagon treatment, and wild type versus knockout) using a two-tailed unpaired Student's *t*-test. **p**, Liver V_{CS} in mice infused with somatostatin, basal insulin and glucagon ($n = 5$). **q**, In vitro oxygen consumption in isolated hepatocytes incubated ± 100 nM glucagon ($n = 113$ wild type – glucagon, 144 knockout – glucagon, 149 wild type + glucagon and 210 knockout + glucagon). **r**, Liver triglyceride content (without glucagon infusion) ($n = 11$). If no statistical comparison is denoted, the groups were not significantly different. Error bars represent s.e.m. All *n* values refer to numbers of mice.



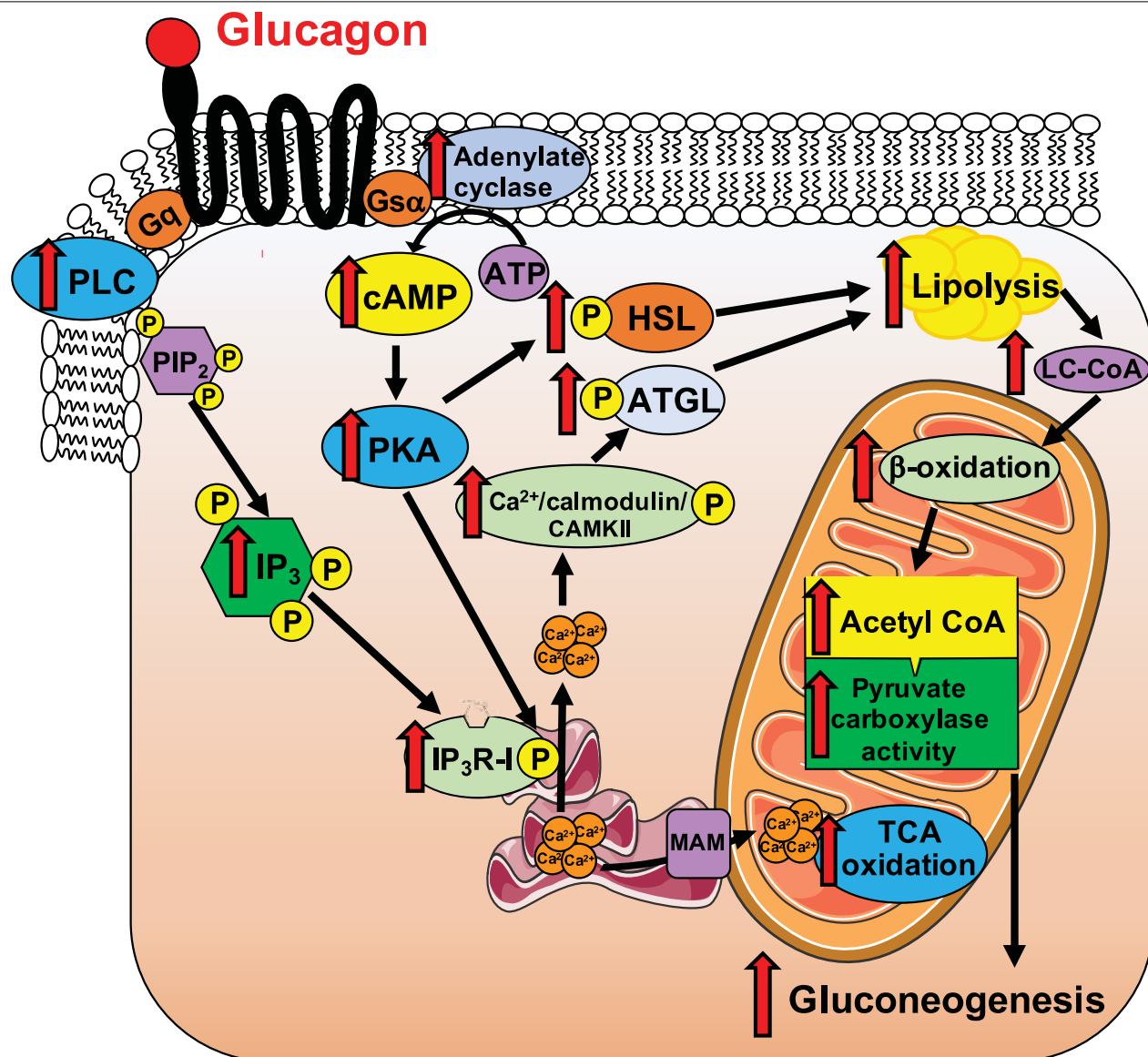
Extended Data Fig. 7 | Chronic increases in mitochondrial oxidation with a ten-day glucagon infusion lead to reversal of nonalcoholic fatty liver disease and improvements in glucose tolerance. **a**, Plasma glucagon concentrations on the last day of infusion ($n=6$ in **a–g**). **b**, Hepatic V_{CS} . In **b–g**, measurements were performed while the glucagon infusion continued. **c**, **d**, V_{PC}/V_{EGP} and V_{PC}/V_{CS} ratios. **e–g**, Hepatic V_{PDH} , V_{FAO} and V_{PK} . **h**, Food intake during the glucagon infusion ($n=8$ control and 10 glucagon), determined twice during the ten-day infusion (on day 4 and day 9) by weighing the food in the cage; the data points are the averages of the two food-intake measurements for each rat. **i**, Body weight after ten days of glucagon or saline infusion ($n=6$). **j**, **k**, Plasma glucose and insulin concentrations ($n=6$) measured two hours after cessation of the glucagon infusion, after a six-hour fast. **l–n**, Liver TAG

($n=6$), DAG ($n=8$) and ceramide ($n=8$) concentrations. In **a–k**, $n=6$. **o**, Hepatic PKCε translocation ($n=6$ control and 7 glucagon). **p–r**, Liver glycogen ($n=6$), acetyl-CoA ($n=8$) and malonyl-CoA content ($n=8$). **s**, **t**, Plasma glucose concentrations and area under the curve (AUC) during an intraperitoneal glucose tolerance test, which began two hours after completing a ten-day continuous infusion of glucagon or saline ($n=6$). In **s**, **u**, $*P<0.05$, $**P<0.01$, $***P<0.001$. Data are mean \pm s.e.m. **u**, **v**, Plasma insulin and insulin area under the curve during the glucose tolerance test. In all panels, error bars represent s.e.m, and groups were compared using a two-tailed unpaired Student's *t*-test. If no statistical comparison is denoted, the groups are not significantly different. All *n* values refer to numbers of rats.



Extended Data Fig. 8 | Chronic glucagon treatment reverses nonalcoholic fatty liver disease and glucose intolerance in wild-type, but not *Insp3r1*-knockout, mice. **a**, Body weight ($n=10$ wild type – glucagon, 11 wild type + glucagon, 8 knockout – glucagon and 8 knockout + glucagon). **b**, **c**, Food and water intake. In **b**–**g**, $n=8$ wild type – glucagon, 9 wild type + glucagon, 7 knockout – glucagon and 8 knockout + glucagon. **d**, Energy expenditure. **e**, **f**, Oxygen consumption and carbon dioxide production. **g**, Activity. **h**, Plasma

NEFA. In **h**–**j**, $n=10$ wild type – glucagon, 11 wild type + glucagon, 8 knockout – glucagon and 8 knockout + glucagon. **i**, **j**, Glucose and insulin area under the curve during an intraperitoneal glucose tolerance test. In all panels, mean \pm s.e.m. is shown. Statistical comparisons were performed using a two-tailed unpaired Student's *t*-test. If no statistical comparison is denoted, the groups were not significantly different. All *n* values refer to numbers of mice.



Extended Data Fig. 9 | Schematic summary. Mechanism by which glucagon acutely stimulates intrahepatic lipolysis, hepatic gluconeogenesis and hepatic mitochondrial oxidation.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Group sizes of 6-10 per group were chosen based on extensive <i>in vivo</i> experience in the lab, to detect moderate to large (>40%) differences with ~40% standard deviations.
Data exclusions	No data were excluded from analysis.
Replication	All attempts at replication were successful. Performing these experiments in the numbers of biological replicates stated, with statistical differences observed where noted, confirms the reproducibility of the findings.
Randomization	Animals were randomly allocated to groups by an investigator who did not participate in any subsequent analysis.
Blinding	Animals were randomly allocated to groups by an investigator who did not participate in any subsequent analysis. All analyses were performed by blinded investigators.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Antibodies

Antibodies used	Antibodies were obtained from Santa Cruz (PC, PEPCK, CAMKII), Cell Signaling (ATGL, pCAMKIV, CAMKIV, pCRTC2, CRTC2, pACC, ACC, pAMPK, AMPK, pHSL, HSL, GAPDH, and -actin), Novus Biologicals (pCAMKII), Abcam (pATGL), LSBio (pInsP3R-I), and BD Transduction Laboratories (InsP3R-II, InsP3R-III, PKCe). The antibody to pATGL was kindly provided by Dr. Hei Sook Sul, while the antibody to total InsP3R-I was custom-made using an epitope against the last 18 amino acids of InsP3R-I.
Validation	The CRTC2 antibody was validated using WT and CRTC2 KO samples provided by Dr. Marc Montminy.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Mice: male IP3R1 KO and WT controls, 10-12 weeks of age Rats: male Sprague-Dawley rats, ~300g, 8-10 weeks of age
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All protocols were approved by the Yale University Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Epigenetic therapy inhibits metastases by disrupting premetastatic niches

<https://doi.org/10.1038/s41586-020-2054-x>

Received: 6 January 2019

Accepted: 28 January 2020

Published online: 26 February 2020

 Check for updates

Zhihao Lu^{1,2,3,17}, Jianling Zou^{2,17}, Shuang Li^{2,17}, Michael J. Topper³, Yong Tao³, Hao Zhang⁴, Xi Jiao², Wenbing Xie³, Xiangqian Kong³, Michelle Vaz³, Huili Li³, Yi Cai³, Limin Xia^{3,5}, Peng Huang³, Kristen Rodgers¹, Beverly Lee¹, Joanne B. Riemer³, Chi-Ping Day⁶, Ray-Whay Chiu Yen³, Ying Cui³, Yujiao Wang², Yanni Wang², Weiqiang Zhang^{1,7}, Hariharan Easwaran³, Alicia Hulbert^{1,8}, KiBem Kim³, Rosalyn A. Jurgens⁹, Stephen C. Yang¹, Richard J. Battafarano¹, Errol L. Bush¹, Stephen R. Broderick¹, Stephen M. Cattaneo¹⁰, Julie R. Brahmer³, Charles M. Rudin¹¹, John Wrangle¹², Yuping Mei^{1,3}, Young J. Kim¹³, Bin Zhang^{14,15}, Ken Kang-Hsin Wang¹⁴, Patrick M. Forde^{3,16}, Joseph B. Margolick⁴, Barry D. Nelkin³, Cynthia A. Zahnow³, Drew M. Pardoll^{3,16}, Franck Housseau^{3,16,18}, Stephen B. Baylin^{3,18}, Lin Shen^{2,18} & Malcolm V. Brock^{1,3,18}

Cancer recurrence after surgery remains an unresolved clinical problem^{1–3}. Myeloid cells derived from bone marrow contribute to the formation of the premetastatic microenvironment, which is required for disseminating tumour cells to engraft distant sites^{4–6}. There are currently no effective interventions that prevent the formation of the premetastatic microenvironment^{6,7}. Here we show that, after surgical removal of primary lung, breast and oesophageal cancers, low-dose adjuvant epigenetic therapy disrupts the premetastatic microenvironment and inhibits both the formation and growth of lung metastases through its selective effect on myeloid-derived suppressor cells (MDSCs). In mouse models of pulmonary metastases, MDSCs are key factors in the formation of the premetastatic microenvironment after resection of primary tumours. Adjuvant epigenetic therapy that uses low-dose DNA methyltransferase and histone deacetylase inhibitors, 5-azacytidine and entinostat, disrupts the premetastatic niche by inhibiting the trafficking of MDSCs through the downregulation of CCR2 and CXCR2, and by promoting MDSC differentiation into a more-interstitial macrophage-like phenotype. A decreased accumulation of MDSCs in the premetastatic lung produces longer periods of disease-free survival and increased overall survival, compared with chemotherapy. Our data demonstrate that, even after removal of the primary tumour, MDSCs contribute to the development of premetastatic niches and settlement of residual tumour cells. A combination of low-dose adjuvant epigenetic modifiers that disrupts this premetastatic microenvironment and inhibits metastases may permit an adjuvant approach to cancer therapy.

Despite advances in traditional cancer treatments and newly developed immunotherapies, cancer recurrence after surgery of non-small-cell lung cancer (NSCLC), oesophageal cancer and breast cancer remains high at 35–76%^{2,8}, 30–66%^{9,10} and 20–66%³, respectively. Most cancer-related mortalities after resection are due to metastases^{1,2}. In early-stage NSCLC in particular, 30–55% of patients die from recurrent metastatic disease after surgery with curative intent^{8,11}, and standard adjuvant chemotherapy confers an absolute 5-year survival benefit of only 3–10%¹¹. Current immune-directed neoadjuvant therapies using immune checkpoint blockade rely on the presence of the primary tumour to generate tumour-antigen-specific T cell responses¹², but in clinical practice the majority of early-stage NSCLC tumours are removed without neoadjuvant intervention^{11,13}. Novel strategies are

needed that decrease tumour recurrence and metastases in the absence of a primary tumour.

MDSCs delineate premetastatic niches

In 2011, we initiated a randomized, phase-II adjuvant clinical trial using low-dose 5-azacytidine and entinostat in patients with stage-I (T1-2aNO) NSCLC. The trial was prematurely terminated after 13 patients enrolled, owing to the requirement for 5-azacytidine to be given in an outpatient clinic. Nevertheless, patients tolerated the therapy well with a low rate of postsurgical recurrence (14.3% versus 33.3%), which suggested that epigenetic therapy may decrease relapses after curative surgery (Extended Data Fig. 1a–c, Supplementary Table 1). A previous

A list of affiliations appears at the end of the paper.

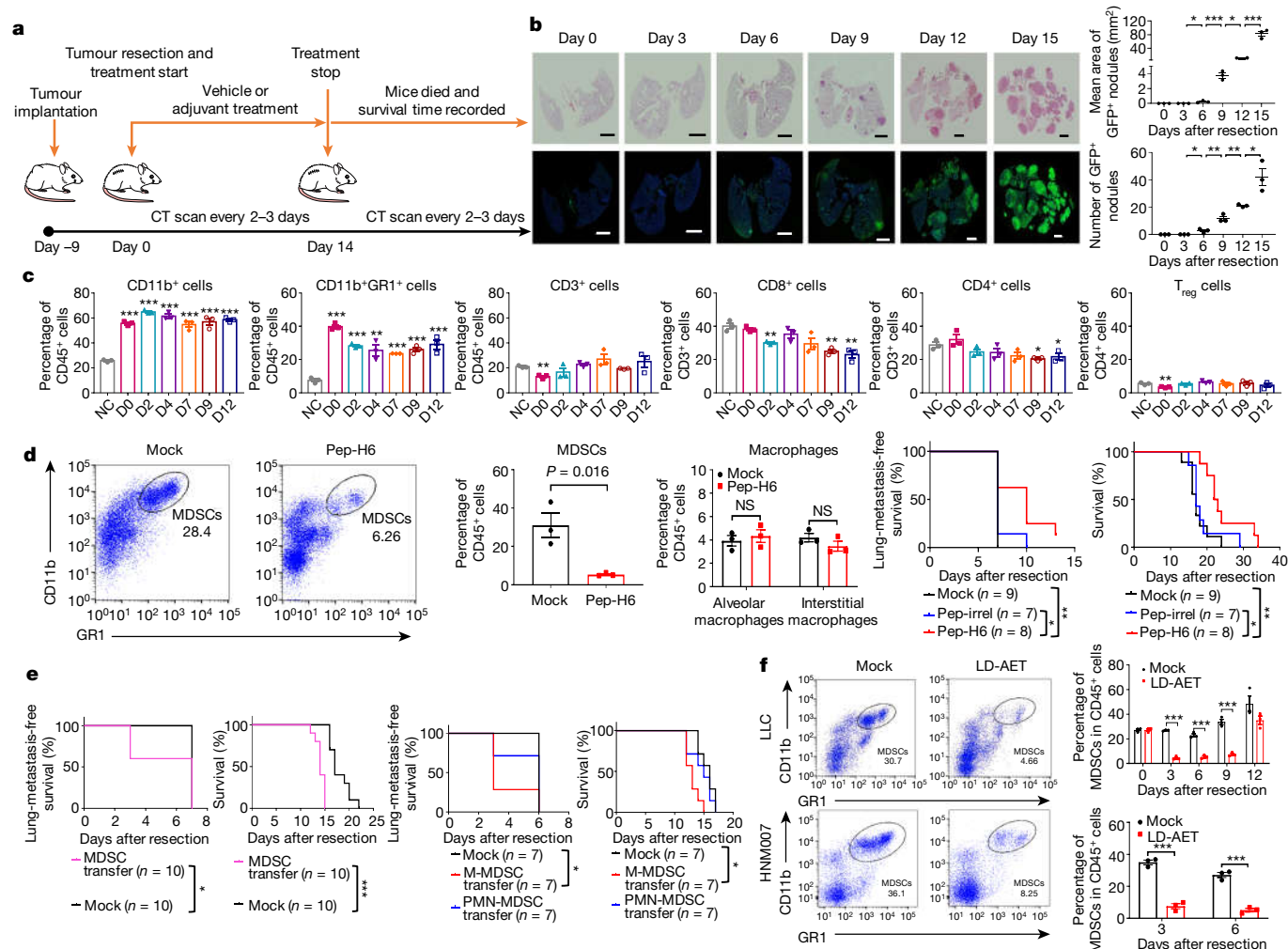


Fig. 1 | Low-dose AET disrupts the lung premetastatic microenvironment by affecting MDSCs. **a**, Timeline of mouse models of metastasis with treatment schedules. CT, computed tomography. **b**, Longitudinal H&E and immunofluorescence staining of lung tissue showed the presence of tumour cells from day 6 in the LLC model. Immunofluorescence staining was performed using GFP (green) antibodies. Merged images contain 4',6-diamidino-2-phenylindole (DAPI) DNA staining, which demarcates cell nuclei (blue). Scale bars, 2 mm. Graph shows the area and numbers of metastatic nodules ($n=3$ mice at each time point). Two-sample, two-sided t -test. **c**, Immune-cell profiles of lungs in the LLC model. Single-cell suspensions from both lungs were analysed by fluorescence-activated cell sorting (FACS) ($n=3$ mice at each time point). NC, negative control (normal lungs from C57BL/6 mice). D, day. Two-sample, two-sided t -test was used in comparison with the negative control. T_{reg} cells, regulatory T cells. **d**, FACS showing lung MDSCs in the LLC model at day 3 were depleted using pep-H6 (leftmost two

panels). Column diagram showing the effect of pep-H6 on the percentages of lung MDSCs and macrophages (middle two panels) at day 3 ($n=3$ mice in each group). Two-sample, two-sided t -test. Kaplan-Meier curves showing disease-free survival and overall survival of LLC mice after MDSC depletion (rightmost two panels). Two-sided log-rank test. Pep-irrel, irrelevant peptidobody, used as control. **e**, Kaplan-Meier curves showing disease-free and overall survival of LLC mice after transfusion of lung MDSCs (5×10^6) (leftmost two panels) and of monocytic (M)-MDSCs (5×10^6) or polymorphonuclear (PMN)-MDSCs (5×10^6) from bone marrow (rightmost two panels). All transfusions were conducted on day 1 and day 4. Two-sided log-rank test. **f**, FACS showing representative effects of low-dose (LD)-AET on lung MDSCs in LLC and HNM007 mice at day 3. Column diagrams showing the effect of low-dose AET on lung MDSCs in LLC and HNM007 mice ($n=3$ mice at each time point). Two-sample, two-sided t -test. Bars show mean \pm s.e.m. NS, not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

study found that combined epigenetic modifiers blunt metastases in a mouse model of aggressive NSCLC with an immune-competent microenvironment¹⁴. Here we reveal a key mechanism of action of adjuvant epigenetic therapy (AET) on premetastatic niches that could underlie these previous findings. Low-dose AET modulates innate immune factors in the lung microenvironment to inhibit tumour recurrence after resection in three syngeneic models of aggressive pulmonary metastasis: Lewis lung carcinoma (LLC), HNM007 oesophageal squamous cell carcinoma (mice of the LLC and HNM007 models have no extrapulmonary metastases) and 4T1 mammary cancer (Fig. 1a, Extended Data Fig. 2a–e).

It is known that the microenvironments of distant organs—the targets of future metastases—are not passive receivers of

circulating tumour cells, but instead are selectively modified by the primary tumour before metastasis^{6,7,15,16}. Settlement of tumour cells at distant sites is dependent on tumour-secreted factors and tumour-shed extracellular vesicles that enable the premetastatic microenvironment to support their colonization^{6,7,15,16}. To study whether low-dose AET affects the dynamics of the premetastatic microenvironment, we used mouse models of highly aggressive, pulmonary metastasis that exhibit pulmonary metastases in 90–100% of mice after resection, with median disease-free survival times of fewer than 14 days (Fig. 1a, b, Extended Data Fig. 2a–d).

In mice of the LLC model (hereafter LLC mice), no tumour-cell infiltration of the lung was detectable by haematoxylin and eosin (H&E) and no green fluorescent protein (GFP)⁺ tumour cells were observed by

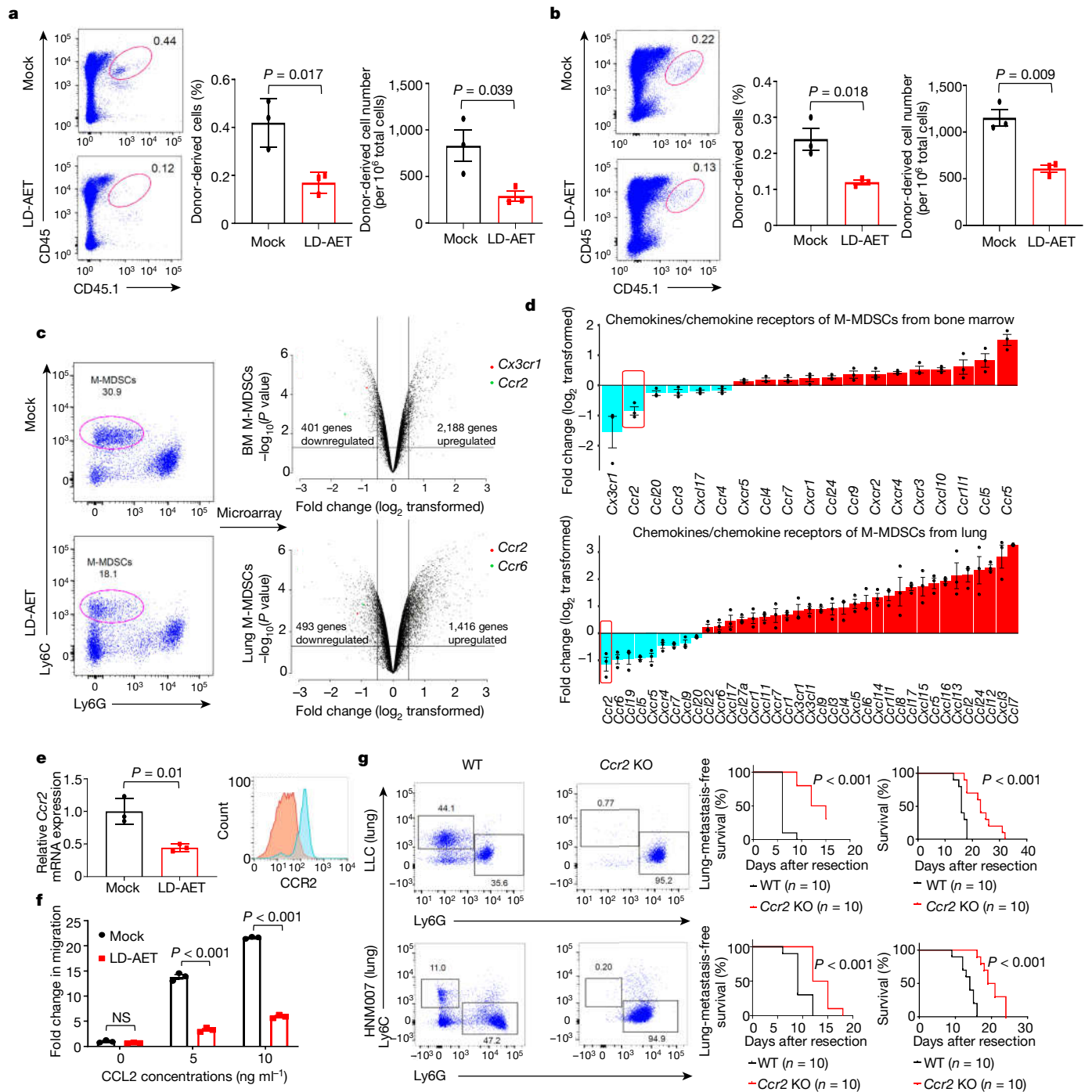


Fig. 2 | Low-dose AET inhibits migration of monocytic MDSCs from the bone marrow to the lung premetastatic microenvironment by downregulating expression of CCR2. **a, b**, The effect of low-dose AET on monocytic MDSCs transferred from CD45.1 to CD45.2 mice (**a**) and the trafficking ability of adoptively transferred monocytic MDSCs from low-dose-AET- or vehicle-treated CD45.1 mice in CD45.2 recipient mice (**b**). FACS and graphs showing percentages and absolute numbers of the donor-derived-cell subset (CD45.1⁺ cells) in both lungs of the recipient LLC mice ($n = 3$ mice per group). **c**, Cell-sorting schema for collecting monocytic MDSCs on day 3 after resection (left). Volcano plots showing differences in RNA expression of monocytic MDSCs from the bone marrow (BM) or lung between low-dose-AET- and vehicle-treated LLC mice (right) ($n = 3$ biological replicates). **d**, Agilent cDNA array expression of chemokines and chemokine receptors in monocytic MDSCs from bone marrow or lungs of low-dose-AET- and vehicle-treated LLC mice ($n = 3$ biological replicates). False-discovery-rate (FDR)-adjusted *P* values of CCR2 are 0.005

(bone marrow) and 0.016 (lung). *Cxcr7* is also known as *Ackr3*. **e**, Effect of low-dose AET on *Ccr2* expression in monocytic MDSCs from bone marrow (BM-M-MDSCs) from LLC mice on day 3, by quantitative PCR (left) ($n = 3$ biological replicates) and FACS (right). In the FACS plot, red represents low-dose AET and blue represents mock treatment. **f**, Transwell migration assay of sorted monocytic MDSCs from bone marrow of low-dose-AET- or vehicle-treated LLC mice (day 3) induced by CCL2 for 60 min. Fold changes are normalized to the migration of the cells in the unstimulated mock group (set at 1) ($n = 3$ biological replicates). **g**, Left, FACS showing representative results of lung MDSCs in *Ccr2*-knockout (KO) and wild-type (WT) LLC and HNM007 mice at day 3. Right, Kaplan-Meier curves showing disease-free survival and overall survival of *Ccr2*-knockout and wild-type LLC (top) and HNM007 (bottom) mice. Two-sided log-rank test. In **a-f**, two-sample, two-sided *t*-tests were used. Bars show mean \pm s.e.m.

immunofluorescence staining until day 6 after resection (Fig. 1b). By contrast, CD11b⁺ cells—including CD11b⁺GRI⁺ cells—are already substantially increased in the lung on the day of resection (day 0) (Fig. 1c). CD11b⁺GRI⁺ cells collected from the lungs on day 3 after primary tumour resection (before visible metastases), inhibited both T cell proliferation and activation in vitro, which suggests that recruited CD11b⁺GRI⁺ cells in the lungs are functional MDSCs (Extended Data Fig. 3a). These MDSCs were the most-increased immune component of all CD11b⁺ cells in tumour-bearing mice immediately before resection. Notably, even after resection, these MDSCs in the LLC and HNM007 models persisted until day 12 in the lung—but not in the liver—as the predominant immune cells, compared to non-tumour-bearing mice (Fig. 1c, Extended Data Fig. 3b, c). We then used pep-H6, a peptibody that selectively targets MDSCs and has a minimal effect on other immune components, to deplete MDSCs¹⁷ (Fig. 1d). Peptibody-mediated depletion of MDSCs resulted in an increase in both disease-free survival and overall survival of the mice (Fig. 1d). By contrast, intravenous injection of MDSCs isolated from the lung of day-3 LLC mice induced metastases sooner and resulted in shorter overall survival times (Fig. 1e), indicating that these MDSCs had an important role in metastasis in this system. Our results indicate that the accumulation of MDSCs in the lungs preceded metastasis development and delineated premetastatic niches. To test the differential role of monocytic (CD11b⁺Ly6C^{high}Ly6G[−]) and polymorphonuclear (CD11b⁺Ly6C^{low}Ly6G⁺) MDSCs in the formation of lung metastases, we transferred monocytic or polymorphonuclear MDSCs from the bone marrow of LLC mice. We found that—compared to vehicle—transferring monocytic MDSCs (but not polymorphonuclear MDSCs) resulted in a higher rate of lung metastasis at day 3 in LLC mice, which suggests that monocytic MDSCs have a predominant role in establishing the premetastatic microenvironment (Fig. 1e).

Low-dose AET impedes migration of MDSCs

It has recently been found that low doses of 5-azacytidine and entinostat are well-tolerated and clinically effective in heavily pretreated patients with NSCLC¹⁸, and that these treatments target MDSCs^{19,20}. Low-dose 5-azacytidine (100 nM) and entinostat (50 nM) in vitro had limited effect on the proliferation of LLC1, HNM007 and 4T1 cells (Extended Data Fig. 4a). Similarly, these doses did not influence the viability or apoptosis of MDSCs sorted from the bone marrow of LLC mice and mice of the HNM007 model (hereafter HNM007 mice) (Extended Data Fig. 4b, c). We determined the in vivo doses of 5-azacytidine and entinostat (0.5 mg per kg body weight per day and 5 mg per kg body weight per day, respectively) that had no effect on primary tumour growth and that did not cause weight loss in immune-compromised mice with LLC and HNM007 tumours (Extended Data Fig. 4d–f). These doses also had a limited effect on the proliferation and apoptosis of CD45.1⁺ donor cells in vivo (Extended Data Fig. 4g, h). Importantly, in our mouse models, these doses decreased MDSCs and niche-promoting molecules in the premetastatic lung (Fig. 1f, Extended Data Fig. 5a–e). Moreover, the percentage of donor CD45.1⁺ MDSCs in the lung is not affected by low-dose AET in the control sham-surgery mice (tumour-naïve recipient mice) (Extended Data Fig. 5f). On the basis of these findings, we hypothesize that low-dose AET can inhibit the accumulation of MDSCs in the lung and prevent the formation of premetastatic niches in our models of pulmonary metastasis.

When CD45.1⁺ monocytic or polymorphonuclear MDSCs (5×10^6 cells each) were adoptively transferred on day 0 to CD45.2 mice, 36 h after transfusion CD45.1⁺ cells decreased by 40–80% in the lungs of recipient mice treated with low-dose AET (Fig. 2a, Extended Data Fig. 6a). When the same numbers of CD45.1⁺ monocytic or polymorphonuclear MDSCs isolated from bone marrow of mock- (day 3) and low-dose AET-treated mice (day 3) were adoptively transferred into CD45.2 mice on day 0, 18 h after transfusion there were significantly fewer CD45.1⁺ cells from mice treated with low-dose AET than from vehicle-treated mice in the lungs

of CD45.2⁺ recipient mice, as expected (Fig. 2b, Extended Data Fig. 6b). These results demonstrate that low-dose AET impedes the migration of MDSCs to the premetastatic microenvironment in the LLC model. Together with our finding that only the transfusion of monocytic (and not of polymorphonuclear) MDSCs increases lung metastases (Fig. 1e), these results showed that—although low-dose AET impairs the migration of both monocytic and polymorphonuclear MDSCs—the role of low-dose AET in targeting the trafficking of monocytic MDSCs may be more important than its targeting of polymorphonuclear MDSCs in our LLC model.

To identify differences in MDSCs from the lung and bone marrow of mock- and low-dose-AET-treated mice on day 3 after resection, we compared the gene expression of monocytic MDSCs (excluding differentiated MHC-II⁺ and F4/80⁺ macrophages) sorted from these two groups in LLC mice (Fig. 2c). Gene-set enrichment analysis (GSEA) of monocytic MDSCs from the lung showed that low-dose AET induced a substantial change in gene sets associated with immune-cell chemotaxis and migration (Extended Data Fig. 6c). CCR2 expression in monocytic MDSCs from both the bone marrow and lung was significantly downregulated in the low-dose-AET group (Fig. 2d). Because CCR2 is a key regulator of monocytic cell migration from the bone marrow to the tumour microenvironment²¹, these data suggest that low-dose AET may affect the trafficking of monocytic MDSCs to the premetastatic lung at least in part by downregulating CCR2.

Quantitative PCR and flow cytometry confirmed that both messenger RNA (mRNA) and protein levels of CCR2 in monocytic MDSCs from bone marrow decreased after low-dose AET (Fig. 2e). Monocytic MDSCs from bone marrow collected on day 3 from LLC mice that were treated with low-dose AET show reduced migration in a transwell assay after induction with CCL2 (Fig. 2f). Both the absolute number and percentage of monocytic MDSCs in the lung premetastatic microenvironment are negligible in *Ccr2*-knockout mice, and differed from wild-type mice (Fig. 2g). Compared to the wild-type C57BL/6 mice, *Ccr2*-knockout mice have a longer disease-free survival and overall survival both in the LLC and HNM007 models (Fig. 2g).

We next tested the mechanism of action of low-dose AET on CCR2 expression in monocytic MDSCs from bone marrow. Database for Annotation, Visualization and Integrated Discovery (DAVID) pathway analysis reveals that the activity of the NF- κ B signalling pathway was significantly downregulated in bone-marrow monocytic MDSCs from LLC mice that have been treated with low-dose AET (Extended Data Fig. 6d). In monocytic MDSCs from bone marrow, low-dose AET resulted in a highly significant reduction in RELB and p52 activation, compared to that found in mock-treated mice (Extended Data Fig. 6e). There was a limited effect on p50 and p65 activation. Furthermore, three days of treatment with BMS-345541 (a highly selective IKK kinase (IKK) allosteric site inhibitor) resulted in decreased expression of CCR2 in monocytic MDSCs from bone marrow in vivo (Extended Data Fig. 6f). Although we cannot rule out a direct effect of low-dose AET on CCR2 expression (as well as other signalling pathways), our findings suggest that low-dose AET treatment may affect—at least in part—the expression of CCR2 in monocytic MDSCs from bone marrow via the modulation of the noncanonical NF- κ B pathway²².

CXCR2 and CXCR1 are known to be important for their role in trafficking polymorphonuclear MDSCs from the bone marrow to the tumour microenvironment^{23,24}. We found that CXCR2 is downregulated in polymorphonuclear MDSCs from both the bone marrow and lung by low-dose AET in the LLC model (Extended Data Fig. 6g). The migration of polymorphonuclear MDSCs from the bone marrow of mice treated with low-dose AET is significantly decreased after induction with CXCL1 in a transwell migration assay (Extended Data Fig. 6h). Thus, low-dose AET may inhibit the trafficking of both monocytic and polymorphonuclear MDSCs from the bone marrow to the premetastatic microenvironment by downregulating CCR2 and CXCR2 expression, respectively.

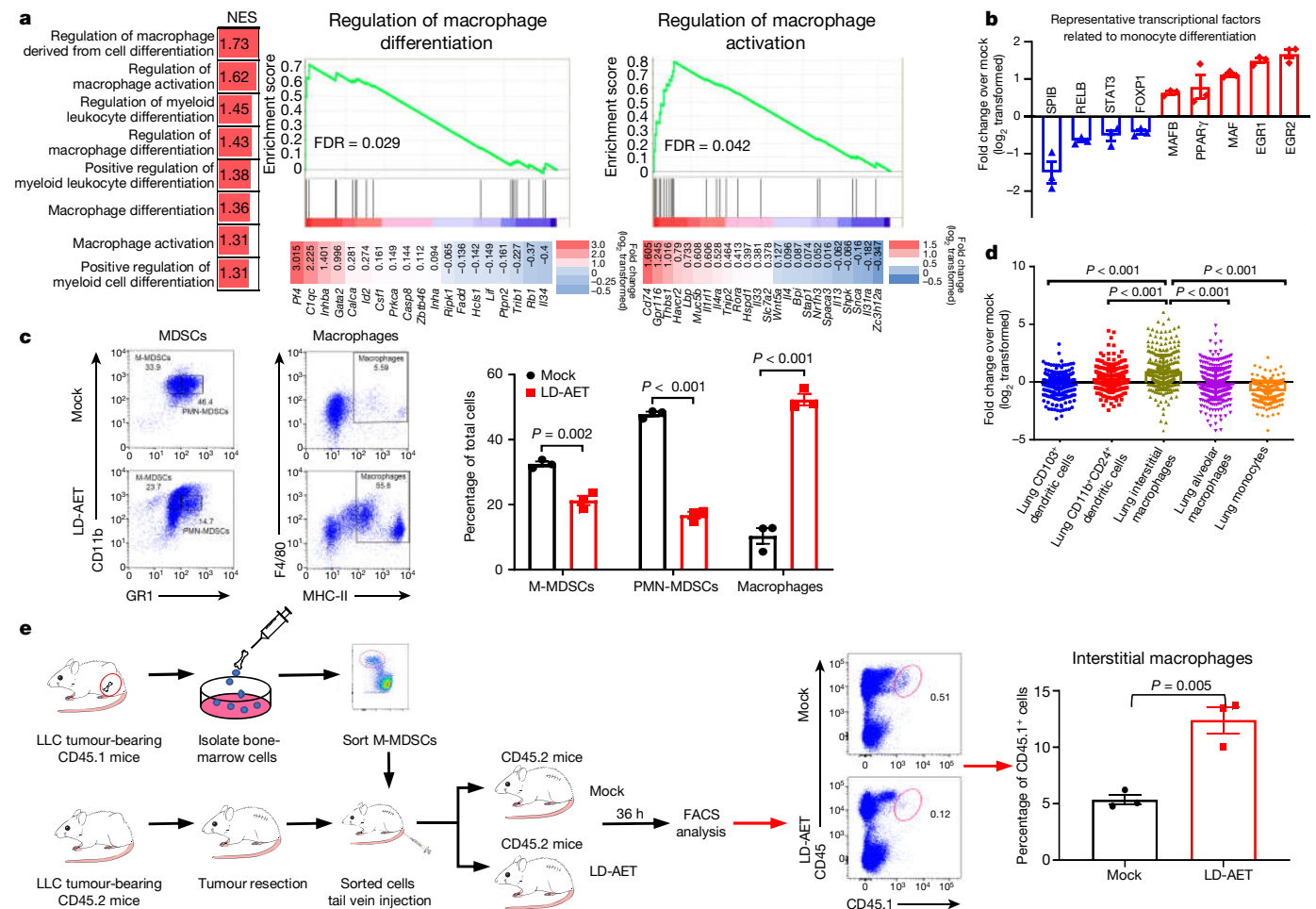


Fig. 3 | Low-dose AET skews monocytic MDSCs towards an interstitial macrophage-like population in the lung premetastatic microenvironment.

a, GSEA analysis revealed that macrophage or myeloid differentiation and activation gene sets were upregulated (left) in monocytic MDSCs from the lungs of LLC mice treated with low-dose AET. Representative upregulated GSEA plots with core-enriched genes (middle and right). NES, normalized enrichment score. Colour gradation is representative of log₂-transformed fold change over mock ($n = 3$ biological replicates). Gene-set enrichment P values, NES values and FDR values reported are calculated with 1,000 permutations in the GSEA software. FDR $q < 0.25$ was deemed significant. *Gpr116* is also known as *Adgrf5*. **b**, Significant changes of representative transcription factors (FDR-adjusted $P < 0.05$) associated with monocytic differentiation ($n = 3$ biological replicates). **c**, FACS showing the differentiation of sorted monocytic MDSCs in vitro. Splenic monocytic MDSCs from LLC mice were cultured for 3 days with tumour-conditioned medium ($n = 3$ biological replicates). **d**, The top 200

significantly upregulated genes (FDR-adjusted $P < 0.05$) were mapped back to the reference ImmGen populations. The following populations from naive mice were used: lung CD103⁺ dendritic cells (ImmGen code, DC_103+11b⁺_Lu); CD11b⁺CD24⁺ lung dendritic cells (DC_103-11b+24+_Lu); lung interstitial macrophages (MF_11c-11b+_Lu); lung alveolar macrophages (MF_Alv_Lu); and lung monocytes (Mo_Lu). The expressions of these 200 genes in 6 populations were transformed by zero-mean normalization. Two-sided Mann–Whitney U -test, $n = 3$ biological replicates. **e**, FACS showing that low-dose AET can skew the differentiation of transferred monocytic MDSCs towards an interstitial macrophage-like phenotype in vivo. Purified 5×10^6 CD45.1⁺ monocytic MDSCs from bone marrow (day 0) were adoptively transferred into CD45.2⁺ recipient mice within 24 h of resection ($n = 3$ in each group) in the LLC model. Mice received low-dose AET or vehicle treatment for 36 h. In **b**, **c** and **e**, two-sample, two-sided t -tests were used. All bars show mean \pm s.e.m.

Skewing monocytic MDSC differentiation

Further query of the GSEA-derived data indicated that gene sets related to macrophage or myeloid differentiation and activation were upregulated in monocytic MDSCs from the lungs of LLC mice treated with low-dose AET (Fig. 3a). Moreover, DAVID pathway analysis reveals that the NF- κ B and PPAR signalling pathways were significantly downregulated and upregulated, respectively (Extended Data Fig. 7a). Consistent with differential regulation of these pathways occurring during the monocyte-to-macrophage differentiation²⁵, there is a preferential increase in transcription factors associated with monocytic differentiation that are mainly macrophage-related (EGR1, EGR2, MAFB, MAF and PPAR γ)^{25,26}, and not in those that are dendritic-cell-related (SPIB, RELB, STAT3 and FOXP1)^{26,27} (Fig. 3b). The related factors have also been validated in vitro and in vivo by quantitative PCR and western

blot analyses (Extended Data Fig. 7b, c). In vitro epigenetic treatment of splenic monocytic MDSCs from mice bearing LLC tumours decreased the percentage and absolute number of monocytic MDSCs significantly, whereas those of macrophages increased significantly (Fig. 3c)—implying that low-dose AET might promote the differentiation of monocytic MDSCs into macrophages.

Characterization using Immunological Genome Project (ImmGen) criteria further defined the resulting populations in the lung premetastatic microenvironment²⁸. Our top 200 upregulated genes in monocytic MDSCs from the lungs of mice treated with low-dose AET map specifically to lung interstitial macrophages (Fig. 3d, Supplementary Table 2). Additionally, flow cytometry shows that lung interstitial macrophages (CD11b⁺CD11c⁺CD64^{high}MHC-II⁺CD24⁺)²⁹ increase in mice treated with low-dose AET (Extended Data Fig. 8a, b). This observation is confirmed by adoptively transferring 5×10^6 CD45.1⁺ monocytic MDSCs

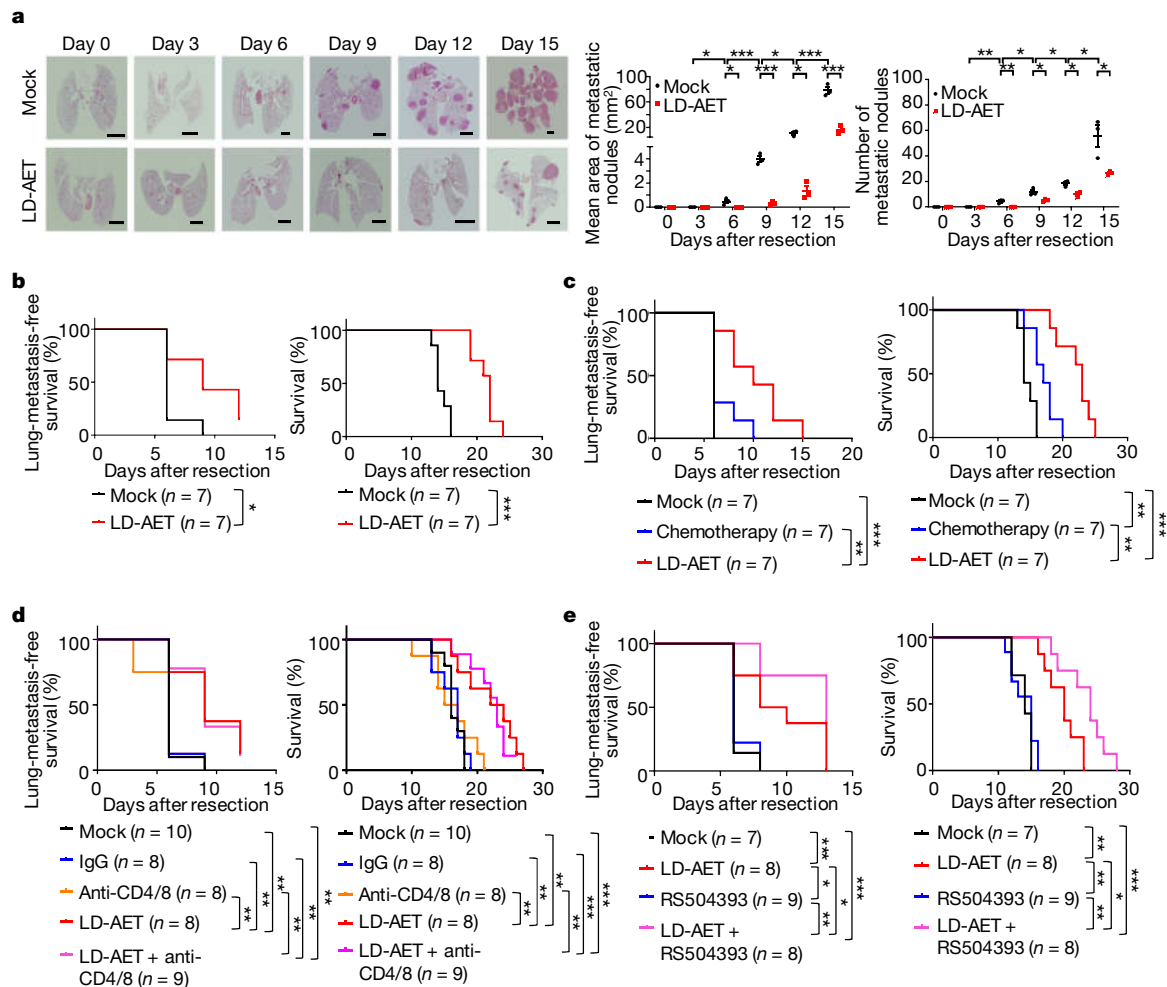


Fig. 4 | Low-dose AET inhibits pulmonary metastases and prolongs overall survival in mouse models, mainly by affecting MDSCs. a, Representative H&E-stained images of lung sections from LLC mice treated with low-dose AET or vehicle at different time points after surgery. Scale bars, 2 mm. Graph shows the area and numbers of metastatic nodules. At each time point, three mice were killed for analysis. For each sample, sections from three levels were analysed. Tumour area was quantified using Aperio Imagescope software. Two-sample, two-sided *t*-test. All bars are mean \pm s.e.m. **b**, Kaplan-Meier curves showing the disease-free survival and overall survival of LLC mice treated with low-dose AET or vehicle. **c**, Kaplan-Meier curves showing the disease-free

survival and overall survival of LLC mice treated with paclitaxel plus cisplatin chemotherapy, low-dose AET or vehicle. **d**, Kaplan-Meier curves showing the disease-free survival and overall survival of mice treated with vehicle, IgG (isotype control), anti-CD4/CD8 (T-cell-depletion antibody), low-dose AET and anti-CD4/CD8 in combination with low-dose AET in the LLC model. **e**, Kaplan-Meier curves showing the disease-free survival and overall survival of mice treated with vehicle, CCR2 antagonist (RS504393) (Sigma), low-dose AET and RS504393 in combination with low-dose AET in the LLC model. In **b**–**e**, two-sided log-rank tests were used. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

into CD45.2⁺ mice. As expected, 36 h after transfer, the CD45.1⁺ monocytic MDSCs differentiated towards a more-interstitial macrophage-like phenotype in the lungs of CD45.2 mice treated with low-dose AET (Fig. 3e, Extended Data Fig. 8c). In summary, low-dose AET may skew monocytic MDSCs towards an interstitial macrophage-like phenotype in the lung premetastatic microenvironment. This acquisition of a MDSC-to-macrophage program is associated with CCR2 down regulation (a known consequence of monocyte-to-macrophage differentiation³⁰), thus also establishing the functional implications of our transcriptional data. Moreover, we found that transferring wild-type monocytic MDSCs rescued metastases in *Ccr2*-knockout mice and resulted in shorter overall survival times, whereas transferring wild-type polymorphonuclear MDSCs did not. By contrast, transferring wild-type monocytic MDSCs from mice treated with low-dose AET brought only limited change to both the lung metastatic rate and overall survival time of *Ccr2*-knockout mice (Extended Data Fig. 8d, e). These findings imply that the decreased accumulation of monocytic MDSCs in the premetastatic niche owing to low-dose AET is mainly dependent on CCR2 signalling.

Low-dose AET increases overall survival

In our mouse models, low-dose AET reduces pulmonary metastases, as shown by H&E staining, gross pathological findings and cone-beam computed tomography imaging (Fig. 4a, Extended Data Fig. 9a–c). In all three mouse models, low-dose AET prolongs disease-free survival and overall survival time (Fig. 4b, Extended Data Fig. 9d). Compared to adjuvant chemotherapy, low-dose AET confers both a longer disease-free and overall survival in the LLC model (Fig. 4c). It has previously been found that epigenetic therapy results in a strong attraction of CD8⁺ T cells to NSCLC tumours and a reversal of the immune-exhaustion profile of these cells¹⁴. However, when we depleted CD4 and CD8 T cells (from day 1 to day 6) in our LLC model, we observed that there was no difference during low-dose AET either in disease-free or overall survival—indicating that the reduction in metastases was independent of T cells (Fig. 4d, Extended Data Fig. 9e). Finally, the combination of low-dose AET and CCR2 antagonists in our LLC and HNM007 models shows synergy in disease-free and overall survival (Fig. 4e, Extended Data Fig. 9f, g).

Our observation that the lung premetastatic niche persists even after resection has translational implications. First, the results stress a potential use of epigenetic treatment as an adjuvant therapy focused on the perturbation of MDSCs. Second, our findings suggest that combining low-dose AET with CCR2 antagonists may be an emerging paradigm to prevent the accumulation of MDSCs in the premetastatic niche, thus inhibiting metastases and extending survival. Third, we provide compelling evidence that if monocytic MDSCs successfully migrate to the lung premetastatic niche, epigenetic modifiers can skew the population to a more-interstitial macrophage-like phenotype, antagonizing their prometastatic functionality in that microenvironment (Extended Data Fig. 10). The post-resection recurrence of early-stage cancer (especially for the tumour types studied here) is a considerable clinical challenge, and effective adjuvant therapies are lacking. Low-dose AET represents a potentially efficacious treatment to use in the absence of manifest primary tumour burden after resection. Our therapeutic paradigm may augment the efficacy of early cancer resection (which still represents the most effective treatment), while robustly inhibiting recurrence. We plan to translate these preclinical findings to a clinical trial in early-stage cancer using low-dose AET and CCR2 antagonists to prevent metastatic recurrence.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2054-x>.

- Mahvi, D. A., Liu, R., Grinstaff, M. W., Colson, Y. L. & Raut, C. P. Local cancer recurrence: the realities, challenges, and opportunities for new therapies. *CA Cancer J. Clin.* **68**, 488–505 (2018).
- Arriagada, R. et al. Long-term results of the international adjuvant lung cancer trial evaluating adjuvant cisplatin-based chemotherapy in resected lung cancer. *J. Clin. Oncol.* **28**, 35–42 (2010).
- Pan, H. et al. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N. Engl. J. Med.* **377**, 1836–1846 (2017).
- Kaplan, R. N. et al. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
- Bonapace, L. et al. Cessation of CCL2 inhibition accelerates breast cancer metastasis by promoting angiogenesis. *Nature* **515**, 130–133 (2014).
- Peinado, H. et al. Pre-metastatic niches: organ-specific homes for metastases. *Nat. Rev. Cancer* **17**, 302–317 (2017).
- Steeg, P. S. Targeting metastasis. *Nat. Rev. Cancer* **16**, 201–218 (2016).
- Pignon, J. P. et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE collaborative group. *J. Clin. Oncol.* **26**, 3552–3559 (2008).
- Allum, W. H., Stenning, S. P., Bancewicz, J., Clark, P. I. & Langley, R. E. Long-term results of a randomized trial of surgery with or without preoperative chemotherapy in esophageal cancer. *J. Clin. Oncol.* **27**, 5062–5067 (2009).
- Shapiro, J. et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol.* **16**, 1090–1098 (2015).
- Kris, M. G. et al. Adjuvant systemic therapy and adjuvant radiation therapy for stage I to IIIA completely resected non-small-cell lung cancers: American Society of Clinical Oncology/Cancer Care Ontario clinical practice guideline update. *J. Clin. Oncol.* **35**, 2960–2974 (2017).
- Forde, P. M. et al. Neoadjuvant PD-1 blockade in resectable lung cancer. *N. Engl. J. Med.* **378**, 1976–1986 (2018).
- Postmus, P. E. et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **28** (Suppl. 4), iv1–iv21 (2017).

- Topper, M. J. et al. Epigenetic therapy ties MYC depletion to reversing immune evasion and treating lung cancer. *Cell* **171**, 1284–1300.e21 (2017).
- Liu, Y. & Cao, X. Characteristics and significance of the pre-metastatic niche. *Cancer Cell* **30**, 668–681 (2016).
- Kitamura, T., Qian, B. Z. & Pollard, J. W. Immune cell promotion of metastasis. *Nat. Rev. Immunol.* **15**, 73–86 (2015).
- Qin, H. et al. Generation of a new therapeutic peptide that depletes myeloid-derived suppressor cells in tumor-bearing mice. *Nat. Med.* **20**, 676–681 (2014).
- Juergens, R. A. et al. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. *Cancer Discov.* **1**, 598–607 (2011).
- Kim, K. et al. Eradication of metastatic mouse cancers resistant to immune checkpoint blockade by suppression of myeloid-derived cells. *Proc. Natl Acad. Sci. USA* **111**, 11774–11779 (2014).
- Youn, J.-I. et al. Epigenetic silencing of retinoblastoma gene regulates pathologic differentiation of myeloid cells in cancer. *Nat. Immunol.* **14**, 211–220 (2013).
- Qian, B.-Z. et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* **475**, 222–225 (2011).
- Sun, S. C. The non-canonical NF- κ B pathway in immunity and inflammation. *Nat. Rev. Immunol.* **17**, 545–558 (2017).
- Highfill, S. L. et al. Disruption of CXCR2-mediated MDSC tumor trafficking enhances anti-PD1 efficacy. *Sci. Transl. Med.* **6**, 237ra67 (2014).
- Di Mitri, D. et al. Tumour-infiltrating Gr-1⁺ myeloid cells antagonize senescence in cancer. *Nature* **515**, 134–137 (2014).
- Saeed, S. et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
- Auffray, C., Sieweke, M. H. & Geissmann, F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu. Rev. Immunol.* **27**, 669–692 (2009).
- Miller, J. C. et al. Deciphering the transcriptional network of the dendritic cell lineage. *Nat. Immunol.* **13**, 888–899 (2012).
- Heng, T. S. & Painter, M. W. The immunological genome project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
- Misharin, A. V., Morales-Nebreda, L., Mutlu, G. M., Budinger, G. R. & Perlman, H. Flow cytometric analysis of macrophages and dendritic cell subsets in the mouse lung. *Am. J. Respir. Cell Mol. Biol.* **49**, 503–510 (2013).
- Phillips, R. J., Lutz, M. & Premack, B. Differential signaling mechanisms regulate expression of CC chemokine receptor-2 during monocyte maturation. *J. Inflamm. (Lond.)* **2**, 14 (2005).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

¹Department of Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. ²Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing, China. ³Department of Oncology, The Johns Hopkins School of Medicine, The Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD, USA. ⁴Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁵State Key Laboratory of Cancer Biology, National Clinical Research Center for Digestive Diseases, Xijing Hospital of Digestive Diseases, Air Force Medical University, Xi'an, China. ⁶Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ⁷Department of Thoracic Surgery, The Seventh Medical Center of PLA General Hospital, Beijing, China. ⁸Department of Surgery, University of Illinois College of Medicine, Chicago, IL, USA. ⁹Division of Medical Oncology, McMaster University, Juravinski Cancer Centre, Hamilton, Ontario, Canada. ¹⁰Department of Surgery, Anne Arundel Medical Center, Annapolis, MD, USA. ¹¹Thoracic Oncology Service, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹²Division of Hematology-Oncology, Medical University of South Carolina, Charleston, SC, USA. ¹³Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University, Nashville, TN, USA. ¹⁴Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University, Baltimore, MD, USA. ¹⁵School of Biomedical Engineering, Dalian University of Technology, Dalian, China. ¹⁶Bloomberg-Kimmel Institute for Cancer Immunotherapy, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹⁷These authors contributed equally: Zhihao Lu, Jianling Zou, Shuang Li. ¹⁸These authors jointly supervised this work: Franck Housseau, Stephen B. Baylin, Lin Shen, Malcolm V. Brock. ✉e-mail: frousseau1@jhmi.edu; sbaylin@jhmi.edu; shenlin@bjmu.edu.cn; mabrock@jhmi.edu

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and cell culture

Lewis lung carcinoma cells (LLC1) and 4T1 cells were obtained from ATCC and cultured in RPMI 1640 medium (GIBCO) supplemented with 10% fetal bovine serum (FBS) (GIBCO). HNM007, a *p53*-null mouse oesophageal squamous cell carcinoma cell line transformed by *Hras*^{G12V}, was provided by S. Singhal and cultured in Dulbecco's modified Eagle's medium (DMEM) (GIBCO) containing FBS at 10% v/v. Cells were maintained in a humidified incubator at 37 °C in the presence of 5% CO₂ and passaged every 2–3 days. Cell lines were routinely tested for mycoplasma and immediately tested upon suspicion. None of the cell lines used in the reported experiments tested positive.

LLC1 and HNM007 cells were kept in RPMI 1640 medium or DMEM with a reduced (3%) serum concentration for 48 h. After that time, supernatants were collected, aliquoted and kept at –80 °C. For MDSC culture, tumour supernatant (3% FBS) containing 10 ng/ml GM-CSF (tumour-conditioned medium) was used.

LLC tissue and HNM007 tissue

The LLC tissue (P3 working stock)³¹, labelled with a *Pol2-Luc/GFP* lentiviral vector, was provided by G. Merlino.

To generate HNM007 tissues, mice were injected subcutaneously in the right flank with 1.0×10^5 viable HNM007 cells in 0.1 ml of PBS and Matrigel (1:1, v/v). On reaching 500–750 mm³, tumours were surgically removed, and lung metastases were monitored periodically by cone-beam computed tomography (CBCT) imaging. This transplantation was referred to as passage zero (P0). The lung nodules were identified by CBCT imaging, resected and then subcutaneously transplanted into other mice (passage one (P1)). The same procedure was repeated eight times (from P2 to P9), and the lung nodules were then collected and subcutaneously transplanted into 50 mice as passage ten (P10). When the P10 tumours reached 500–750 mm³, they were collected into 100 tubes and frozen in liquid nitrogen as working stocks. All the HNM007 tumours in our studies were expanded from the P10 working stock.

Mice

Female C57BL/6 mice and BALB/c mice were purchased from Charles River Laboratories. Female mice congenic in mouse CD45 at the Ly5 locus (B.6SJL-*Ptprca*^a *Peptc*^b/BoyJ Ly5.1) were obtained from Jackson Laboratory. Female NOD/SCID/g-chain knockout (NSG) mice, bred and housed at the Johns Hopkins Animal Care Facility, were used. B6.129S4-*Ccr2*^{tm1lf}/J mice were provided by S. A. McGrath-Morrow. All mice were maintained in pathogen-free conditions and used for experiments at age of 6–8 weeks. The Guide for the Care and Use of Laboratory Animals published by the National Institutes of Health (NIH publication 96-01, revised 1996) was followed, and all protocols were approved by the Johns Hopkins University Animal Care and Use Committee.

Lung spontaneous metastasis model

For preclinical studies, a vial of P3 LLC tumour was expanded subcutaneously in 10 C57BL/6 mice (equals P4). The expanded tumours were resected at 500–750 mm³ and transplanted subcutaneously into the required number of mice for the actual drug study (P5). In the HNM007 model, a vial of P10 HNM007 tumour was expanded subcutaneously in 10 C57BL/6 mice (equals P11). The expanded tumours were resected at 500–750 mm³ and transplanted subcutaneously into the required number of mice for the actual drug study (P12). For the 4T1 model, 2×10^5 4T1 cells were injected into the BALB/c mouse mammary fat pad using a tuberculin syringe. For all models of spontaneous lung metastasis, the primary tumours were surgically removed at 500–750

mm³, and the mice were randomized into treatment groups. For the sham surgery study, tumour-naïve mice underwent a 2.0-cm skin incision and closure in right flank. Mice were treated as follows: azacitidine 0.5 mg/kg (PBS vehicle) subcutaneously injected and daily for 14 days. entinostat 5.0 mg/kg (LLC and HNM007 models) and 2.5 mg/kg (4T1 model) (1% dimethyl sulfoxide (DMSO) in PBS vehicle) intraperitoneally injected daily for 14 days. In vivo, the CCR2 inhibitor, RS504393 (Sigma) 2 mg/kg (PBS vehicle) or RS102895 (Sigma) 2 mg/kg (PBS vehicle) was injected intraperitoneally daily for 14 days. In vivo, the NF-κB inhibitor BMS-345541 (Selleck) 30 mg/kg/d or 75 mg/kg/d (3% Tween 80 and sterile water) was orally administered daily for 3 days. The adjuvant chemotherapy regimen was paclitaxel 20 mg/kg/week injected intraperitoneally plus cisplatin 3 mg/kg, injected intraperitoneally twice a week for 2 weeks. The size of the subcutaneous tumours was measured manually and calculated by $V(\text{mm}^3) = 0.5 \times L \times W^2$, in which *L* is length and *W* is width, in millimetres. The metastasis or recurrence was monitored by CBCT imaging. The time from primary-tumour resection to metastasis detection by CBCT imaging and to death was defined as the disease-free survival and overall survival period, respectively.

Xenograft studies in NSG mice

LLC tissue (P5) and HNM007 tissue (P12) were transplanted subcutaneously into the flanks of mice. Drug treatments were started 6–8 days after transplantation, when palpable tumours could be discerned. Treatment was continued for the entire duration of the study and mice were killed before tumour volumes exceeded 2,000 mm³.

CD4⁺ and CD8⁺ T cell depletion experiment

CD4 antibody (BioXcell, clone GK1.5) and CD8 antibody (BioXcell, clone 116-13.1)-mediated CD4⁺ and CD8⁺ T cell depletion in LLC mice was initiated within 24 h of primary-tumour resection. Mice were injected intraperitoneally with 500 µg of CD4 and CD8 antibodies at day 1 and day 4, respectively, after resection. As controls, mouse IgG2a isotype control (BioXcell, clone C1.18.4) and rat IgG2b isotype control (BioXcell, clone LTF-2) were injected into the control mice. CD4⁺ and CD8⁺ T cell depletion was verified by FACS analysis of peripheral blood cells.

MDSC depletion experiment

Synthetic, complementary double-stranded oligonucleotides encoding H6 peptide (TIK), and an irrelevant control peptibody (pep-irrel) (D1) were provided by L. W. Kwak¹⁷. The recombinant peptibodies used in all in vivo studies were produced by Kempbio following an established protocol¹⁷. After LLC tumour resection, groups of C57BL/6 mice were injected via their tail veins with 50 µg of peptibody (TIK) once per day from day 1 to day 14. Control mice received pep-irrel or PBS. MDSC depletion was verified by assessment of the MDSC population in the premetastatic lung.

CBCT guided systems

To monitor the lung metastases, the mice were subjected to CBCT imaging (laboratory of K.K.-H.W.) at different time points after primary tumour resection. The standard procedure for CBCT imaging has previously been described³².

Drug reagents

Azacitidine (Sigma) was dissolved in distilled water at 500 µM (in vitro) and 5 mg/ml (in vivo). Entinostat (MS-275) provided by P. Ordentlich was dissolved in DMSO to concentrations of 500 µM (in vitro) and 1 mg/ml (in vivo). Both azacitidine and entinostat were aliquoted and stored at –80 °C and diluted to needed working concentrations before use. BMS-345541 (Selleck) was formulated as a 2 mg/ml solution in 3% Tween 80, water and stored at 4 °C. RS504393 (Sigma) and RS102895 (Sigma) were first dissolved in DMSO and stored at –80 °C. The stock solution was dissolved in normal saline at 2 mg/ml for intraperitoneal injection. All the stored reagents were for single use only.

Cell viability assays (CCK8 assay)

Equal numbers of viable cells were plated in 96-well plates at the following cell seeding densities per well: LLC1 cells (2×10^3), HNM007 cells (2×10^3), 4T1 cells (2×10^3), monocytic MDSCs (1×10^4) and polymorphonuclear MDSCs (1×10^4). Cells were incubated with 100 μ l drug-supplemented medium, treated with DMSO (vehicle) at 0.1% or the following drug concentrations standardized to 0.1% DMSO final concentration. For LLC1 cells, HNM007 cells and 4T1 cells, the treatment regimens were: azacitidine, 25 nM, 50 nM, 100 nM, 200 nM and 400 nM; entinostat, 25 nM, 50 nM, 100 nM, 200 nM and 400 nM; combined treatment, entinostat 50 nM plus azacitidine 25 nM, 50 nM, 100 nM, 200 nM and 400 nM. After incubation for 24 h, 48 h or 72 h, cell viability was measured using a Cell Counting Kit-8 (CCK8) assay (Dojindo) according to the manufacturer's instructions. The optical density at 450 nm (OD_{450nm}) was measured using a multiwell plate reader (Micro-plate Reader; Bio-Rad).

Apoptosis assay

Apoptotic assays in cultured cells or transferred cells were performed using the Annexin V-FITC apoptosis detection kit (BD Pharmingen). Assays were performed according to the manufacturer's protocol.

T cell suppression assay

MDSCs were isolated using the mouse MDSC isolation kit, according to the manufacturer's protocol (Miltenyi Biotec), or by flow cytometry. CD8⁺ T cells were isolated from the spleens of tumour-bearing mice by magnetic separation as previously described³³, and then labelled with 5 μ M carboxyfluorescein succinimidyl ester (CFSE) (Invitrogen). The CFSE-labelled T cells were incubated with CD11b⁺Ly6C^{low}Ly6G⁺ or CD11b⁺Ly6C^{high}Ly6G⁻ cells isolated from the lung premetastatic niche (day 3) at different ratios in a 96-well plate cultured with CD3/CD28 beads (GIBCO) at 37 °C with 5% CO₂. After 72 h, the cells and supernatant were collected. Cells were stained with anti-CD8 antibody so that CD8⁺ T cells could be specifically gated and examined, and the CFSE fluorescence intensity of the CD8⁺ T cells was determined by flow cytometry. IFN γ concentrations in the supernatant were determined by enzyme-linking immunosorbent assay (ELISA) (R&D systems) according to the manufacturer's instructions.

In vitro transwell migration assay

Monocytic or polymorphonuclear MDSCs isolated from the bone marrow by FACS were incubated in tumour-conditioned medium in the upper chamber of transwell inserts (5- μ m pore for monocytic MDSCs or 3- μ m pore for polymorphonuclear MDSCs) with CCL2 (5 ng/ml or 10 ng/ml) or CXCL1 (20 ng/ml or 50 ng/ml) in the lower chamber. Transmigrated monocytic or polymorphonuclear MDSCs were enumerated following a 60-min or 120-min incubation. Fold changes were normalized to the migration of the cells in the unstimulated mock group (set at 1).

FACS profiling and sorting of immune cells

FACS profiling and sorting of immune cells were performed after both lungs were collected from LLC mice, HNM007 mice and mice of the 4T1 model (hereafter 4T1 mice) in mock-treated or low-dose-AET-treated groups. Tissue was then digested for 30 min at 37 °C in digestion buffer (RPMI, FBS (5%), collagenase type 1 (Sigma-Aldrich, 0.2%), collagenase type 2 (Sigma-Aldrich, 0.2%) and DNase I (Roche, 50 U/ml)), minced and strained through a 40- μ m cell strainer to obtain a single-cell suspension¹⁴. Perfused livers were cut into small fragments and incubated (37 °C, 250 rpm for 30 min) with 5 ml digestion buffer (5% FBS, 0.5 mg/ml collagenase VIII (Sigma-Aldrich) and 0.1 mg/ml DNase I in PBS). This was followed by 3 cycles of washing with PBS at 400 rpm from which the supernatant was taken, omitting the parenchymal cell pellet. Spleens were mechanically dissociated, and red blood cells were lysed in 1 \times lysing buffer (BD Bioscience). Femurs and tibias from mice were dissected,

and the bone marrow was flushed with RPMI 1640 medium. The cells were filtered through a 70- μ m cell strainer. Blood was obtained by tail vein puncture or by heart puncture after exposing the organ. Blood was collected into tubes containing 1.0 ml of PBS with 2 mM EDTA. Red blood cells were then lysed with lysing buffer and the cell pellet was washed twice in PBS. Cells were counted and then blocked with rat monoclonal anti-CD16/CD32 (Fc block antibody) in PBS for 30 min at 4 °C. Cells were then stained with antibodies. For intracellular antigens, cells were fixed and permeabilized in fixation/permeabilization buffer (eBioscience) for 30 min at 4 °C, washed and stained with intracellular antibodies for 30 min at 4 °C. Information about all the antibodies used is provided in Supplementary Table 3.

Cell transfer experiments

For adoptive transfer experiments, MDSCs were sorted from the bone marrow of CD45.1⁺ LLC mice by FACS. A total of 5×10^6 monocytic or polymorphonuclear MDSCs were transferred into a CD45.2 mouse via tail-vein injection within 24 h of resection.

Immunofluorescence analysis

Immediately after perfusion, mice lungs were embedded in optimal cutting temperature compound, snap-frozen and stored at -80 °C until analysed. Five-micrometre cryosections were cut, air-dried, acetone-fixed and blocked with 10% normal goat serum (30 min, Sigma-Aldrich). To detect MDSCs, slides were stained with rat anti-GR1 antibody (1:500, Biolegend). To detect T cells, slides were stained with rabbit anti-CD4 antibody (1:200, Abcam) and rat anti-CD8a antibody (1:100, eBioscience). To detect GFP signal, slides were stained with rabbit anti-GFP (1:200, Abcam). Sections were washed in PBS 3 times before adding goat anti-rabbit Alexa Fluor 488 and goat anti-rat TRITC secondary antibodies in blocking solution for 45 min at room temperature. All sections were mounted in ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific) and imaged with a Nikon Confocal Microscope C1 and analysed with EZ-C1 software (Nikon).

Gene-expression analysis

For genome-wide gene expression analysis, we used gene-expression arrays at the Sidney Kimmel Comprehensive Cancer Center Microarray Core at the Johns Hopkins University. MDSCs were isolated from bone marrow or lungs by FACS. RNA preparations were obtained from monocytic MDSCs from lung or bone marrow pooled from 6–8 mice per sample (low-dose AET group) or from 1–2 mice per sample (vehicle-treatment group). After total cellular RNA was extracted using the TRIzol method (Life Technologies), RNA concentration was determined using the NanoDrop machine and software (Thermo Fisher Scientific). Around 400 ng of total RNA was used to generate cDNA with the QuantiTect Reverse Transcription Kit (Quanta Biosciences). Transcriptomic profiles were obtained using Agilent 4 \times 44K mouse Gene Expression v.2 arrays following the manufacturer's instructions. Microarray data were analysed with the R package limma as described in the manual. In brief, background signals were corrected using the normexp method (with offset = 50). Then, normalization within arrays and between arrays was performed using the losses and Aquantile methods. The differential gene expression was defined as log₂-transformed fold change > absolute (0.5) and an FDR adjusted $P < 0.05$. Ranked lists of log₂-transformed fold change were analysed using GSEA by the Broad-Institute-developed data packages³⁴. Significantly enriched gene sets were defined using an FDR cut-off of <0.25. P values were defined as <0.05 when comparing treated conditions versus controls.

DAVID analysis of median absolute deviation derived genes

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted using the DAVID Bioinformatics resources database^{35,36}. Only categories that were below the FDR adjusted P value of 0.05 were considered.

Quantitative real-time PCR

Quantitative real-time PCR was performed with SYBR Green I detection chemistry (Bio-Rad Laboratories) using the Applied Biosystems 7500 Fast Real-Time PCR System and its software. β -Actin was used as a reference gene. The specific primers used for quantitative real-time PCR are listed in the Supplementary Table 4. The $\Delta\Delta C_t$ method was used to calculate relative expression levels.

Immunoblotting

Protein was extracted by RIPA buffer containing protease and phosphatase inhibitors. Six-to-ten per cent Bis-Tris protein gels were equally loaded with 30 μ g protein, electrophoresed at 110 V, and electrotransferred to PVDF membranes. Membranes were blocked with 10% milk in TBST and immunoblotted with the following antibodies: rabbit polyclonal anti-MMP-9 (Abcam, 1:1,000), rabbit monoclonal anti-TGF β (Abcam, 1:1,000), rabbit polyclonal anti-ARG1 (Abcam, 1:1,000), rabbit monoclonal anti-VEGF-A (Abcam, 1:1,000), rabbit monoclonal anti-S100A8 (Abcam, 1:1,000), rabbit monoclonal anti-TNF (Abcam, 1:1,000), rabbit monoclonal anti-IL-6 (Cell Signaling Technology, 1:1,000), rabbit polyclonal anti-DNMT1 (Cell Signaling Technology, 1:1,000), rabbit monoclonal anti-EGR1 (Abcam, 1:1,000), rabbit monoclonal anti-EGR2 (Abcam, 1:1,000), rabbit monoclonal anti-PPAR γ (Cell Signaling Technology, 1:1,000), mouse monoclonal anti-MAF-B (Santa Cruz Biotechnology, 1:100), rabbit monoclonal anti-p50 (Cell Signaling Technology, 1:1,000), rabbit monoclonal anti-p52 (Cell Signaling Technology, 1:1,000), rabbit monoclonal anti-RELB (Cell Signaling Technology, 1:1,000), rabbit monoclonal anti-p65 (Cell Signaling Technology, 1:1,000), mouse monoclonal anti- β -actin (Sigma Aldrich, 1:10,000). The loading control antibodies (anti- β -actin) in all cases were applied. Information about all the antibodies used is provided in Supplementary Table 3.

NF- κ B DNA-binding capability assay

To measure NF- κ B activation, the TransAM NF- κ B Family Transcription Factor Assay Kit (43296, Active Motif) was used according to manufacturer's protocol. In brief, bone-marrow-derived monocytic MDSCs from LLC mice treated with vehicle or low-dose AET were isolated, and nuclear extracts were prepared in lysis buffer AM2. Nuclear lysates were incubated with oligonucleotides containing the NF- κ B-binding consensus sequence, and specific antibodies were used to detect the different subunits within the bound complexes. Quantification was performed via colorimetric readout of absorbance at 450 nm.

H&E staining and imaging

For histological analysis, lungs were fixed in 10% formalin overnight, and subsequently transferred into 70% ethanol, embedded in paraffin according to standard protocols. Sections (5 μ m) were stained with H&E and viewed under the Nikon Eclipse NiE microscope (Nikon Instruments). Images from the whole slide were acquired by Nikon NIS Element software. For each sample, sections from three levels were analysed. The number and area of metastatic sites were quantitated using Aperio Imagescope software.

Adjuvant epigenetic treatment trial

The J1037 (NCT01207726) study was a randomized phase-II study that compared the low-dose AET (5-azacytidine plus entinostat) with standard of care (observation) in patients with stage I (T1-2aNO) NSCLC after primary tumour resection. The J1037 study was performed in full accordance with the guidelines for Good Clinical Practice and the Declaration of Helsinki, and all patients gave written informed consent. Protocol approval was obtained from the Johns Hopkins Hospital and Anne Arundel Medical Center Ethics Committee. An independent data monitoring committee reviewed the safety data. The patients from the adjuvant epigenetic treatment group received the combination

of azacitidine at 40 mg/m² on days 1–5 and 8–10 with entinostat at a 7-mg fixed dose on days 3 and 10 of each 28-day cycle. The primary end point was the effect of 5-azacytidine plus entinostat on the hazard of 3 years of progression-free survival in patients with resected stage I non-small-cell lung cancer. Finally, 13 patients were enrolled in the trial. Owing to the difficulty in enrolling patients, the trial was prematurely terminated on 1 May 2015.

Statistical analysis

Flow and imaging data were collected using FlowJo Version 10.0.7, or Summit Version 5.4 (Beckman Coulter). All the experiments were performed in biological and technical triplicates. Values reported in figures are expressed as the standard error of the mean, unless otherwise indicated. Depending on the type of experiment, data were tested using two-sample, two-sided *t*-test, two-sided log-rank test, one-way analysis of variance (ANOVA), hypergeometric test or Mann–Whitney *U*-test. *P* values of < 0.05 were considered significant. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Statistical analyses were performed with GraphPad Prism 7.0 (GraphPad Software) or R version 3.6.1 (<https://www.r-project.org>).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated are included in the Article and in its Supplementary Information. Source Data for Figs. 1–4, Extended Data Figs. 1, 3–9 are provided with the paper. Gene-expression data that support the findings of this study have been deposited in the Gene Expression Omnibus under accession number GSE124539. All data are also available from the corresponding authors on reasonable request.

- Day, C. P., Carter, J., Bonomi, C., Hollingshead, M. & Merlino, G. Preclinical therapeutic response of residual metastatic disease is distinct from its primary tumor of origin. *Int. J. Cancer* **130**, 190–199 (2012).
- Yu, J. et al. Systematic study of target localization for bioluminescence tomography guided radiation therapy. *Med. Phys.* **43**, 2619 (2016).
- Dieckmann, D., Plottner, H., Berchtold, S., Berger, T. & Schuler, G. Ex vivo isolation and characterization of CD4⁺CD25⁺ T cells with regulatory properties from human blood. *J. Exp. Med.* **193**, 1303–1310 (2001).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
- Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).

Acknowledgements This work was supported by grants from the Brockman Foundation, the Skalka-Kronsberg family as well as the Banks Family Foundation, Bermuda. Research funding was provided by the Van Andel Institute through the Van Andel Institute–Stand Up To Cancer Epigenetics Dream Team. Stand Up To Cancer is a division of the Entertainment Industry Foundation, administered by AACR. We acknowledge G. Merlino for providing the LLC tissue (P3 working stock); S. Singhal, who provided HNM007, a p53-null mouse oesophageal squamous cell carcinoma cell line transformed by *Hras*^{G12V}; S. A. McGrath-Morrow, who provided B6.129S4 *Ccr2*^{tm1mif}/J mice; L. W. Kwak, who provided synthetic, complementary double-stranded oligonucleotides encoding H6 peptide (TIK), and an irrelevant control peptide (Irr-pep) (D1); A. Tam and R. L. Blosser for their help with flow cytometry; P. Ordentlich from Syndax Pharmaceuticals for providing entinostat; S. Zhou for advice and consultation; L. Bois and J. Murphy for expert technical help; Y. Lai for histological analysis with H&E staining; and W. Zhu for data analysis.

Author contributions Z.L., J.Z., S.L. and M.V.B. conceptualized, designed and performed the experiments and wrote the manuscript. S.B.B., L.S., F.H. and D.M.P. designed experiments and wrote the manuscript. M.J.T., Y.T. and M.V. helped with the design the experiments and reviewing the manuscript. K.R. and B.L. helped with the sample collection. C.-P.D. provided LLC tissue in the study. H.Z., W.X., X.K., H.L., X.J., Yanni Wang, Yujiao Wang, R.-W.C.Y., W.Z., Y. Cai, H.E.Y. Cui, L.X., A.H., J.B.R. and Y.M. performed experiments and assisted in acquisition of data. C.M.R. and R.A.J. designed the clinical trial. K.K., S.C.Y., R.J.B., E.L.B., S.R.B., S.M.C., J.R.B., J.W., Y.J.K. and P.M.F. enrolled patients and assisted in acquisition of data. P.H. helped with the statistical analyses. B.Z. and K.K.-H.W. helped with the CBCT imaging. C.A.Z. assisted with the mouse experiments. J.B.M. and B.D.N. assisted with discussions and reviewing the manuscript.

Article

Competing interests Z.L. and M.V.B. have patent applications on epigenetic therapy. J.R.B., C.M.R., P.M.F. and R.A.J. serve on the advisory boards of Bristol-Meyers Squibb (BMS) and AstraZeneca. J.R.B., P.M.F. and R.A.J. are on the advisory board of Merck and receive grant support from BMS. J.R.B., C.M.R. and R.A.J. are on the advisory boards of Amgen and Genentech/Roche. Y.J.K., P.M.F. and R.A.J. serve on the advisory board of Novartis. C.M.R. and J.R.B. are on the advisory board of Celgene. J.R.B. and P.M.F. serve on the advisory boards of Eli Lilly and Jansen. C.M.R. and P.M.F. are on the advisory board of AbbVie. R.A.J. and P.M.F. serve on the advisory board of Boehringer Ingelheim. R.A.J. and Y.L.K. are on the advisory board of Takeda. R.A.J. is on the advisory board of Pfizer. S.R.B. is a consultant to BMS. J.W. is cofounder and chief medical officer of Precision Genetics. J.R.B. serves on the advisory board of Syndax, and receives honoraria from Roche/Genentech. C.M.R. is a consultant to and/or advisory board member for Ascentage, Daiichi Sankyo, Ipsen, Loxo, Pharmamar, Vavotek, Bridge Medicines and Harpoon Therapeutics. Y.J.K. serves on the advisory boards of Dracen, Aduro and Sanofi.

P.M.F. is a consultant to and/or advisory board member for EMD Serono and Inivata, and receives grant support from Corvus, Kyowa and Novartis. S.B.B. serves on the advisory boards of Mirati Therapeutics, MDxHealth and Aminex Therapeutics. M.V.B., Y.M., J.R.B., P.M.F. and R.A.J. receive grant support from AstraZeneca. R.A.J. and J.R.B. receive grant support from Merck. All other authors declare no competing interests.

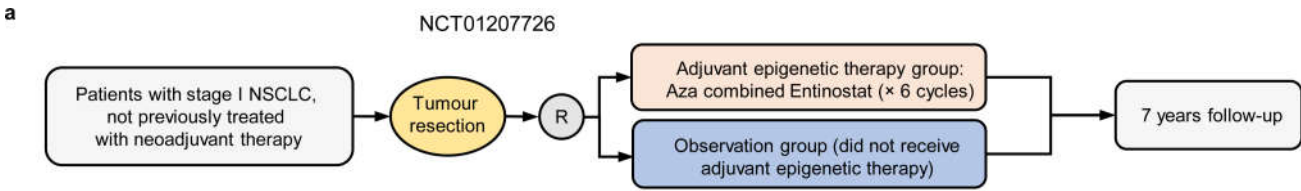
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2054-x>.

Correspondence and requests for materials should be addressed to F.H., S.B.B., L.S. or M.V.B.

Peer review information *Nature* thanks Michele De Palma, Maria Figueroa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

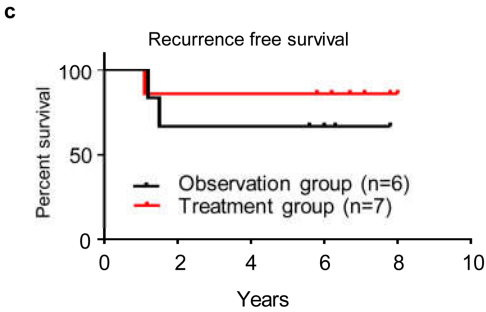
Reprints and permissions information is available at <http://www.nature.com/reprints>.



*Agents administered:
Aza administered at 40 mg/m², SC on days 1-5, 8-10 of each cycle (q28d)
Entinostat administered at 7 mg, PO on d3 and d10 of each cycle (q28d)
R, randomization

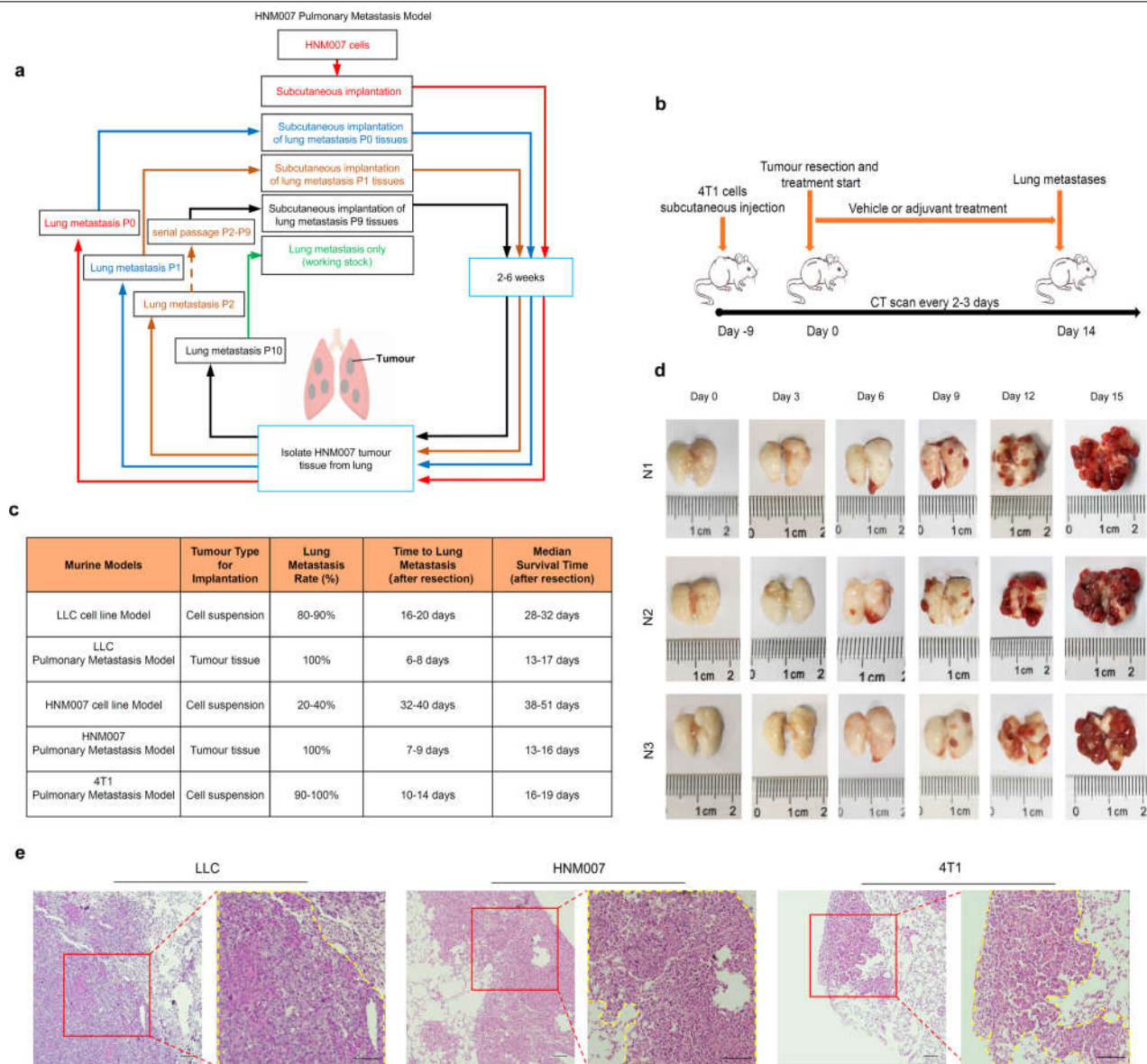
b

	Observation n (%)	Adjuvant epigenetic therapy n (%)	Total n
Recurrence	2 (33.3%)	1 (14.3%)	3
No-recurrence	4 (66.7%)	6 (85.7%)	10
Total	6 (100%)	7 (100%)	13



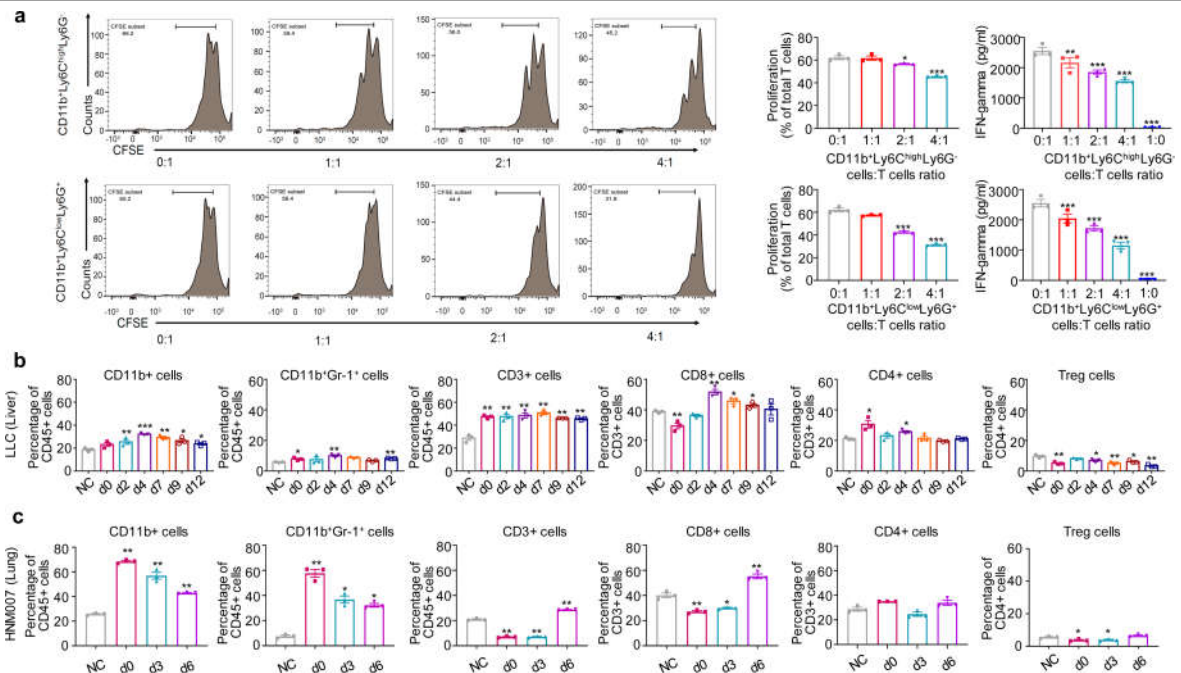
Extended Data Fig. 1 | Efficacy of low-dose AET on cancer recurrence in patients with stage I (T1-2aN0) NSCLC in a phase-II clinical trial. a. Schema for a randomized phase-II clinical trial of AET in patients with stage I (T1-2aN0)

NSCLC (NCT01207726). **b.** Postsurgical recurrence rates in the observation and AET groups. **c.** Kaplan–Meier curves for disease-free survival in the observation and AET groups. $P=0.50$ by two-sided log-rank test.



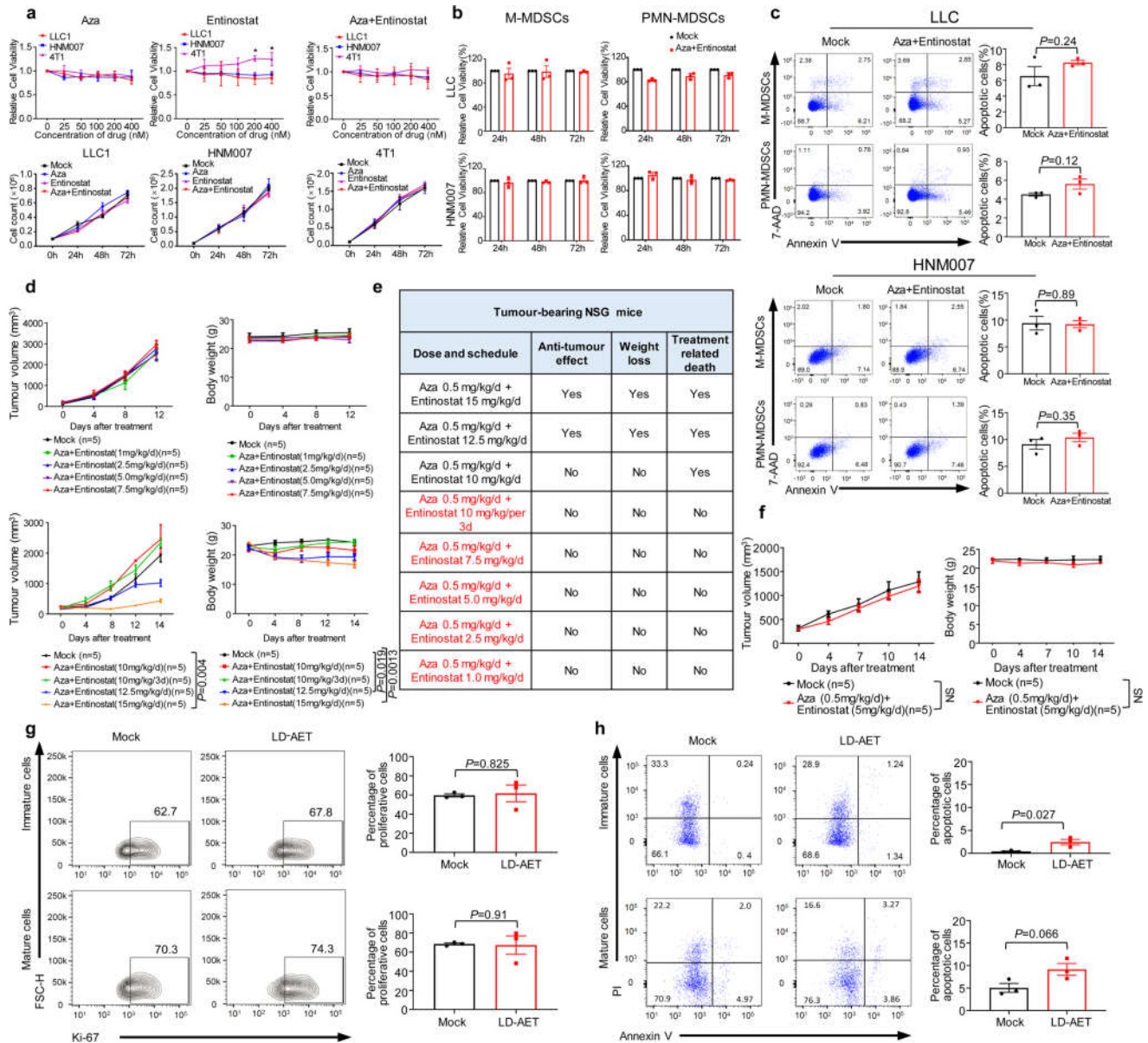
Extended Data Fig. 2 | Schema outlining the establishment and characteristics of the mouse models of pulmonary metastasis. a, Schema for establishing the highly aggressive HNMO07 model. Pulmonary metastases were collected and serially subcutaneously implanted in the right flanks of mice for 10 passages. **b**, Schema for establishing the 4T1 model. **c**, Characteristics of mouse cell line (nonselective) and spontaneous mouse tissue models of pulmonary metastasis (metastases collected selectively from serial pulmonary metastases to produce a solely pulmonary-metastatic phenotype). **d**, Longitudinal gross pathological photographs of bilateral

pulmonary metastases during the natural history of the LLC model in C57BL/6 mice from day 0 to day 15 after surgery. N1, N2 and N3 depict the experiment performed in triplicate. Two mice were killed at each time point from day 0 to day 15 ($n = 36$); data for 18 mice are shown here as representative photomicrographs. **e**, H&E staining of pulmonary metastases in LLC, HNMO07 and 4T1 mice. The histology of LLC (day 9), HNMO07 (day 12) and 4T1 (day 12) pulmonary metastases were confirmed by a pathologist. Scale bars, 100 μ m. Representative data were repeated at least three times with similar results.



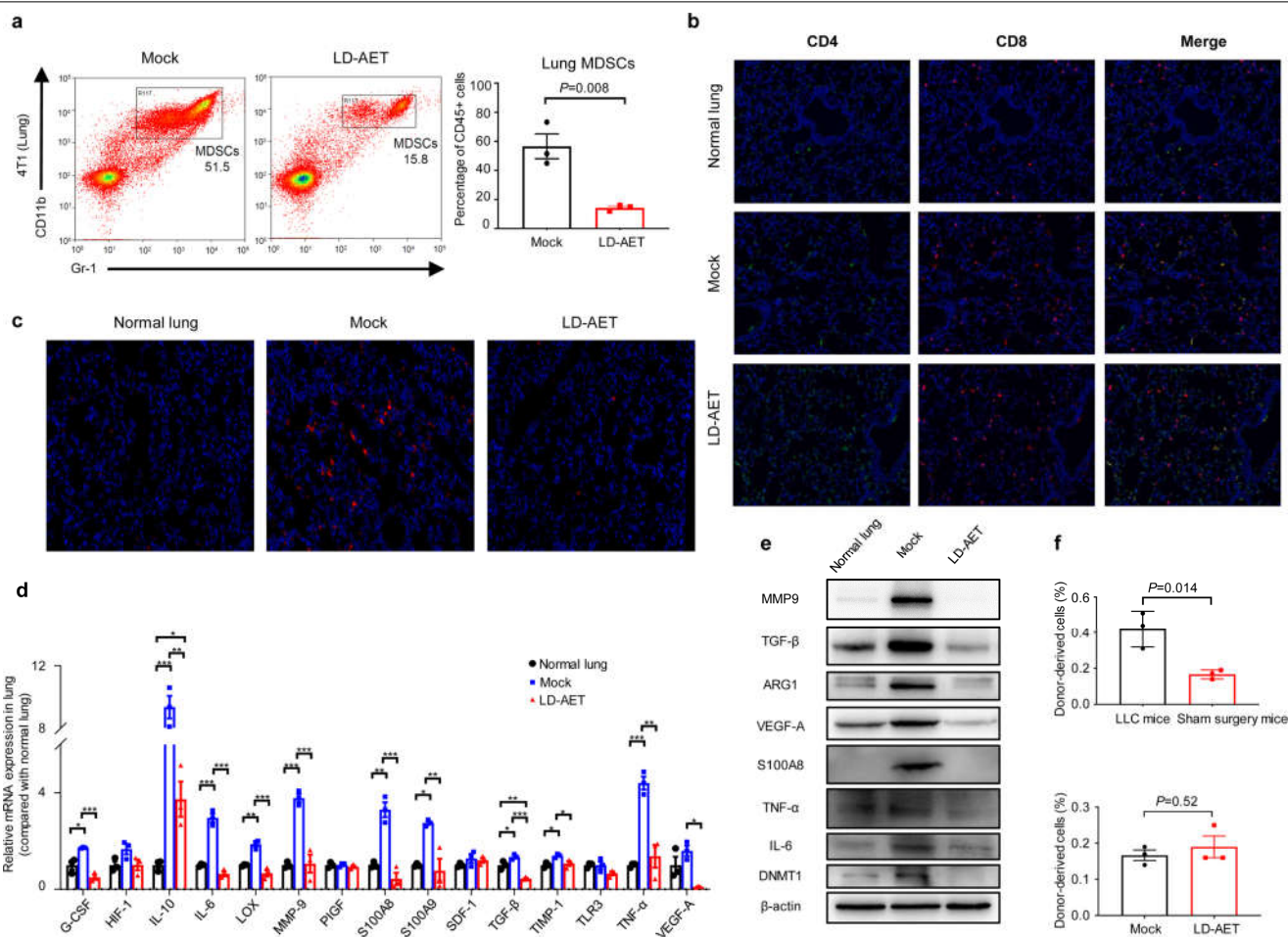
Extended Data Fig. 3 | CD11b⁺GR1⁺ cells persist as the predominant immune cells even after resection in the lung premetastatic microenvironment as functional MDSCs. **a**, In LLC mice, lung CD11b⁺Ly6C^{high}Ly6G⁻ and CD11b⁺Ly6C^{low}Ly6G⁺ cells collected at 72 h after resection both have suppressive activity in vitro against CD8a T cells. Freshly isolated CD11b⁺Ly6C^{high}Ly6G⁻ or CD11b⁺Ly6C^{low}Ly6G⁺ cells from both lungs at day 3 after resection were cocultured with CD8a T cells for 72 h at different ratios (0:1, 1:1, 2:1, 4:1 and 1:0). T cell proliferation and IFN γ concentrations in the supernatant were measured by FACS (left) and ELISA (right), respectively ($n=3$ biological replicates). Representative data were repeated at least three times with similar

results. Two-sample, two-sided t -test was used in the comparison with mock (CD8a T cells alone). **b**, Immune-cell profiles of liver in LLC mice. Single-cell suspensions from the entire liver were analysed by FACS ($n=3$ mice per time point) at different time points after surgery. NC, negative control (normal liver from C57BL/6 mice). **c**, Immune-cell profiles of both lungs in HNM007 mice at different time points after surgery. Single-cell suspensions from both lungs were analysed by FACS ($n=3$ mice per time point). NC, negative control (normal lungs from C57BL/6 mice). In **b**, **c**, a two-sample, two-sided t -test was used in comparison with the negative control. All bars show mean \pm s.e.m. * $P<0.05$, ** $P<0.01$, *** $P<0.001$.



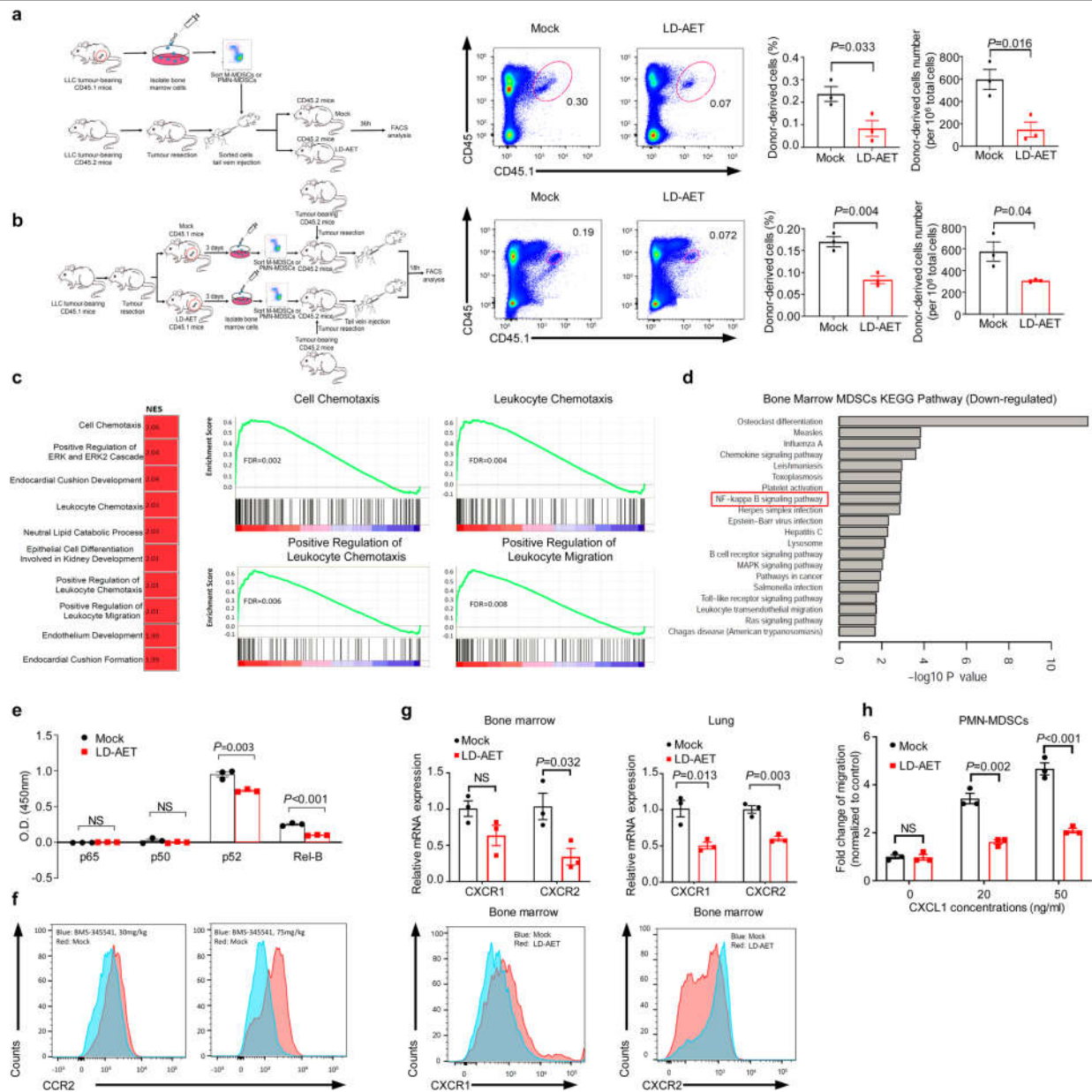
Extended Data Fig. 4 | Consideration of combined AET dosing on the basis of its effect on mouse models. **a**, Top, effect of different dosages of epigenetic modifiers on the viability of LLC1, HNM007 and 4T1 cells in vitro (72 h, Cell Counting Kit-8). Graphs show the mean of 3 independent experiments; two-sample, two-sided *t*-tests compared with mock. Bottom, effect of low-dose 5-azacytidine (100 nM) plus entinostat (50 nM) on the proliferation of LLC1, HNM007 and 4T1 cells in vitro. A total of 1×10^5 viable cells was plated per well. Cells were collected at 24, 48 and 72 h and counted using a cell counter (Bio-Rad) after Trypan blue exclusion. Graphs show the mean of 3 independent experiments; significance at 72 h was determined by one-way ANOVA followed by Tukey's test for multiple comparisons. **b**, The effect of low-dose 5-azacytidine (100 nM) plus entinostat (50 nM) on the viability of MDSCs from bone marrow (day 3) of LLC mice (top) and HNM007 mice (bottom) in vitro (Cell Counting Kit-8). Graphs show the mean of 3 independent experiments. **c**, The effect of low-dose 5-azacytidine (100 nM) plus entinostat (50 nM) on the apoptosis of MDSCs from bone marrow (day 3) of LLC and HNM007 mice in vitro. Cell apoptosis was measured by FACS at 48 h. The bottom right quadrant (annexin-V⁺/7-AAD⁻) and top right quadrant (annexin-V⁺/7-AAD⁺) represent early and late apoptotic cells, respectively. Graphs show the

percentage of total apoptosis (early and late apoptosis) in mock and treatment groups ($n=3$ biological replicates). **d**, Tumour growth and body weight of NSG mice bearing LLC tissue that were treated with different doses of entinostat plus 5-azacytidine. Significance at day 12 (top) and day 14 (bottom) was determined by one-way ANOVA followed by Tukey's test for multiple comparisons. **e**, Summary table of tumour growth, body weight and treatment-related death of NSG mice bearing LLC tissue. Regimens in red indicate dosages with no effect on tumour growth, weight loss or treatment-related death. **f**, Tumour growth and body weight of NSG mice bearing HNM007 tissue that were treated with 5-azacytidine at $0.5 \text{ mg kg}^{-1} \text{ d}^{-1}$ plus entinostat at $5 \text{ mg kg}^{-1} \text{ d}^{-1}$ or vehicle. Significance at day 14 was determined by two-sample, two-sided *t*-test. **g, h**, The effect of low-dose AET on the proliferation (**g**) and apoptosis (**h**) of donor-derived CD45.1⁺ MDSCs from bone marrow in CD45.2 LLC mice. Proliferation and apoptosis of immature (MHC-II⁻) and mature (MHC-II⁺) CD45.1⁺ cells were measured by FACS at 36 h after transfusion (day 2). Graphs in **g** show the percentage of Ki67⁺ cells ($n=3$ mice per group). Graphs in **h** show the percentage of total apoptosis (early and late apoptosis) in mock-treated and low-dose-AET groups ($n=3$ mice per group). In **b, c, g, h**, two-sample, two-sided *t*-tests were used. All bars show mean \pm s.e.m. * $P < 0.05$.



Extended Data Fig. 5 | Low-dose AET disrupts the lung premetastatic microenvironment, mainly by affecting MDSCs. **a**, The effect of low-dose AET (5-azacytidine $0.5 \text{ mg kg}^{-1} \text{ d}^{-1}$ plus entinostat $2.5 \text{ mg kg}^{-1} \text{ d}^{-1}$) on MDSCs from the lung at day 3 after resection in 4T1 mice ($n = 3$ mice per group). **b**, **c**, Immunofluorescence staining of CD4⁺ and CD8⁺ T cells (**b**) or GR1⁺ cells (**c**) from the lung premetastatic microenvironment (day 3) in LLC mice with or without low-dose AET. Negative control was normal lungs from tumour-free C57BL/6 mice. Immunofluorescence staining was performed using CD4 (green) and CD8 (red) antibodies, or GR1 (red) antibodies. Merged images contain DAPI staining for cell nuclei (blue). Original magnification 20 \times . Representative data were repeated at least three times with similar results. **d**, **e**, The mRNA (**d**) and protein (**e**) levels of representative molecular factors known to promote premetastatic microenvironment formation from both lungs of normal mice, and mock- or low-dose-AET-treated LLC mice (day 3) were measured by

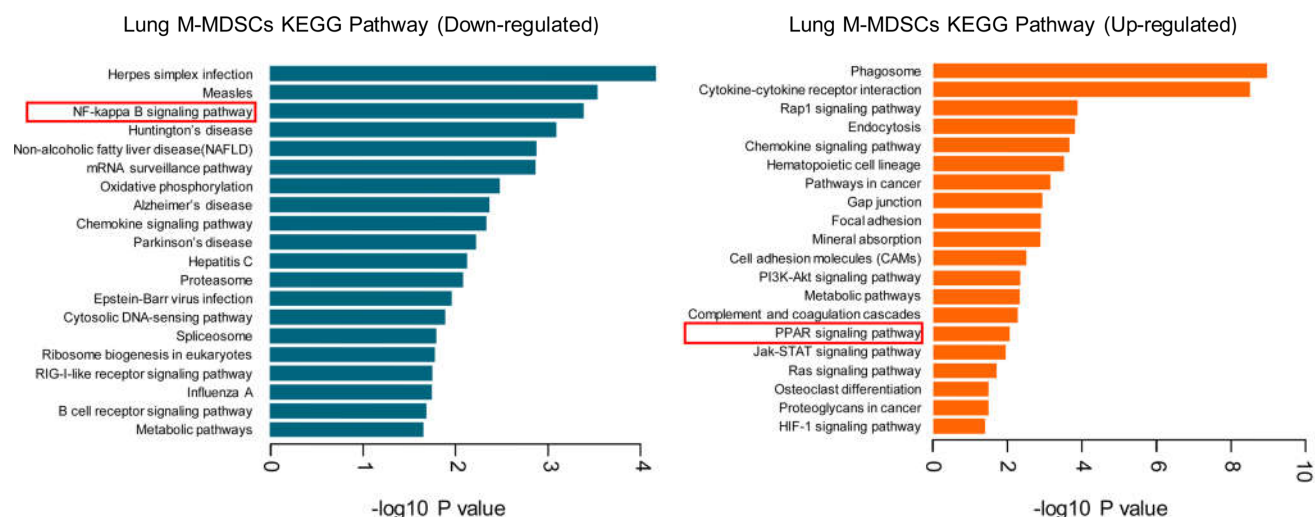
quantitative PCR and western blot. Two-sample, two-sided *t*-test for quantitative PCR experiments ($n = 3$ biological replicates). For gel source data, see Supplementary Fig. 1. All the experiments were performed in triplicate and similar results were obtained. **f**, Top, graph showing the percentages of donor-derived cell subsets (CD45.1⁺ MDSC cells) in the lungs of LLC mice or of sham-surgery mice (tumour-naïve recipient mice) 36 h after surgery. Bottom, graph showing the percentages of donor-derived cell subsets (CD45.1⁺ MDSC cells) in the lungs of low-dose-AET- or vehicle-treated sham-surgery mice (tumour-naïve recipient mice) 36 h after surgery. Purified 5×10^6 MDSCs from bone marrow of CD45.1 mice bearing LLC tumours (day 0) were adoptively transferred into CD45.2 recipient mice in the sham-surgery tumour-naïve model or LLC model ($n = 3$ mice per group). In **a**, **f**, two-sample, two-sided *t*-tests were used. All bars show mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



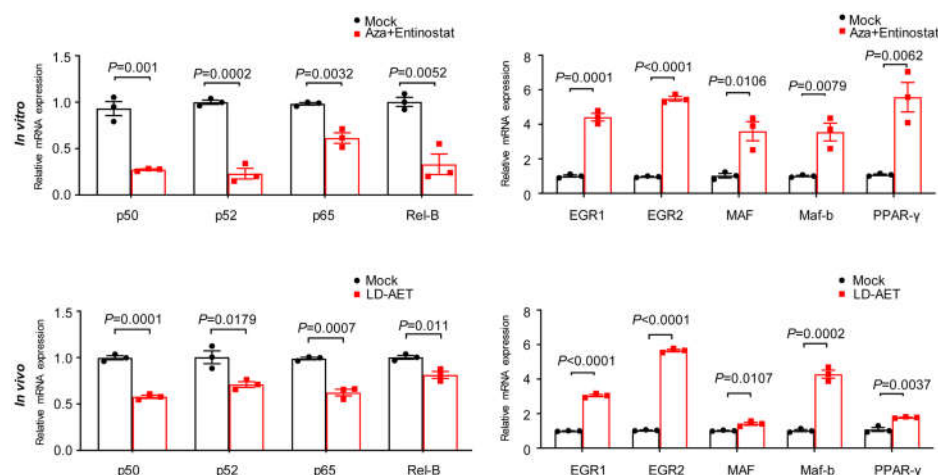
Extended Data Fig. 6 | Low-dose AET induces substantial changes in immune-cell chemotaxis and migration in MDSCs in LLC mice. **a**, Left, schema showing the effect of low-dose AET on monocytic or polymorphonuclear MDSCs transferred from CD45.1 to CD45.2 C57BL/6 mice in the LLC model. CD45.1⁺ cells (transferred polymorphonuclear MDSCs) were identified in the lungs of the recipient mice and analysed by flow cytometry (right, $n = 3$ mice per group). **b**, Left, schema showing trafficking ability of adoptively transferred monocytic or polymorphonuclear MDSCs from low-dose AET-treated or untreated CD45.1 mice in the LLC model. CD45.1⁺ cells (transferred polymorphonuclear MDSCs) were identified in the lungs of the recipient CD45.2 mice and analysed by flow cytometry at 18 h after transfer (right, $n = 3$ mice per group). **c**, Left, top 10 upregulated gene sets from GSEA of lung monocytic MDSCs after 72 h of treatment with low-dose AET. Middle and right, representative upregulated GSEA plots of immune-cell chemotaxis and migration. Colour gradation is representative of \log_2 -transformed fold change over mock ($n = 3$ biologically replicates). Gene-set enrichment P values, NES values and FDR values reported are calculated with 1,000 permutations in the GSEA software. FDR $q < 0.25$ was deemed significant. **d**, DAVID analyses of significantly downregulated genes using the KEGG Gene Ontology in MDSCs from bone marrow of LLC mice, treated or untreated with low-dose AET. Top 20

downregulated pathways are presented ($n = 3$ biologically replicates). Hypergeometric test (FDR-adjusted $P < 0.05$). **e**, Low-dose AET significantly decreases the nuclear activation of p52 and RELB ($OD_{450\text{nm}}$ (mock versus low-dose AET), p52: 0.95 ± 0.035 versus 0.721 ± 0.011 , $P = 0.0034$; RELB, 0.251 ± 0.012 versus 0.1 ± 0.003 , $P = 0.0002$), but not p50 and p65 in monocytic MDSCs from bone marrow of LLC mice in vivo. Nuclear lysates were incubated with oligonucleotides containing the NF- κ B-binding consensus sequence, and specific antibodies were used to detect the different subunits within the bound complexes ($n = 3$ biological replicates). **f**, FACS shows the effect of 30 mg $\text{kg}^{-1} \text{d}^{-1}$ and 75 mg $\text{kg}^{-1} \text{d}^{-1}$ of BMS-345541 (a highly selective IKK kinase inhibitor) on CCR2 expression in monocytic MDSCs from bone marrow of LLC mice on day 3 after surgery. The experiments were performed in triplicate, and similar results were obtained. **g**, CXCR1 and CXCR2 expression of polymorphonuclear MDSCs collected on day 3 from the bone marrow or lung detected by quantitative PCR (top) and FACS (bottom) in LLC mice treated with vehicle or with 72 h of low-dose AET ($n = 3$ biological replicates). **h**, Transwell migration assay of sorted polymorphonuclear MDSCs from bone marrow of low-dose AET- (72 h) or vehicle-treated LLC mice, induced by CXCL1 (20 ng ml^{-1} and 50 ng ml^{-1}) for 120 min ($n = 3$ biological replicates). In **a**, **b**, **e**, **g**, **h**, two-sample, two-sided t -tests were used. All bars show mean \pm s.e.m.

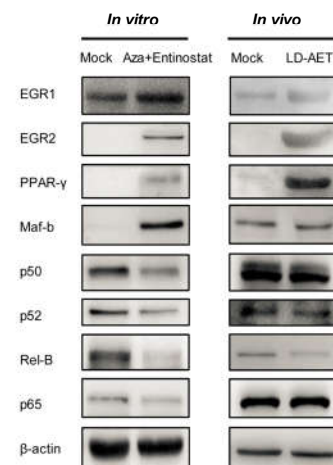
a



b

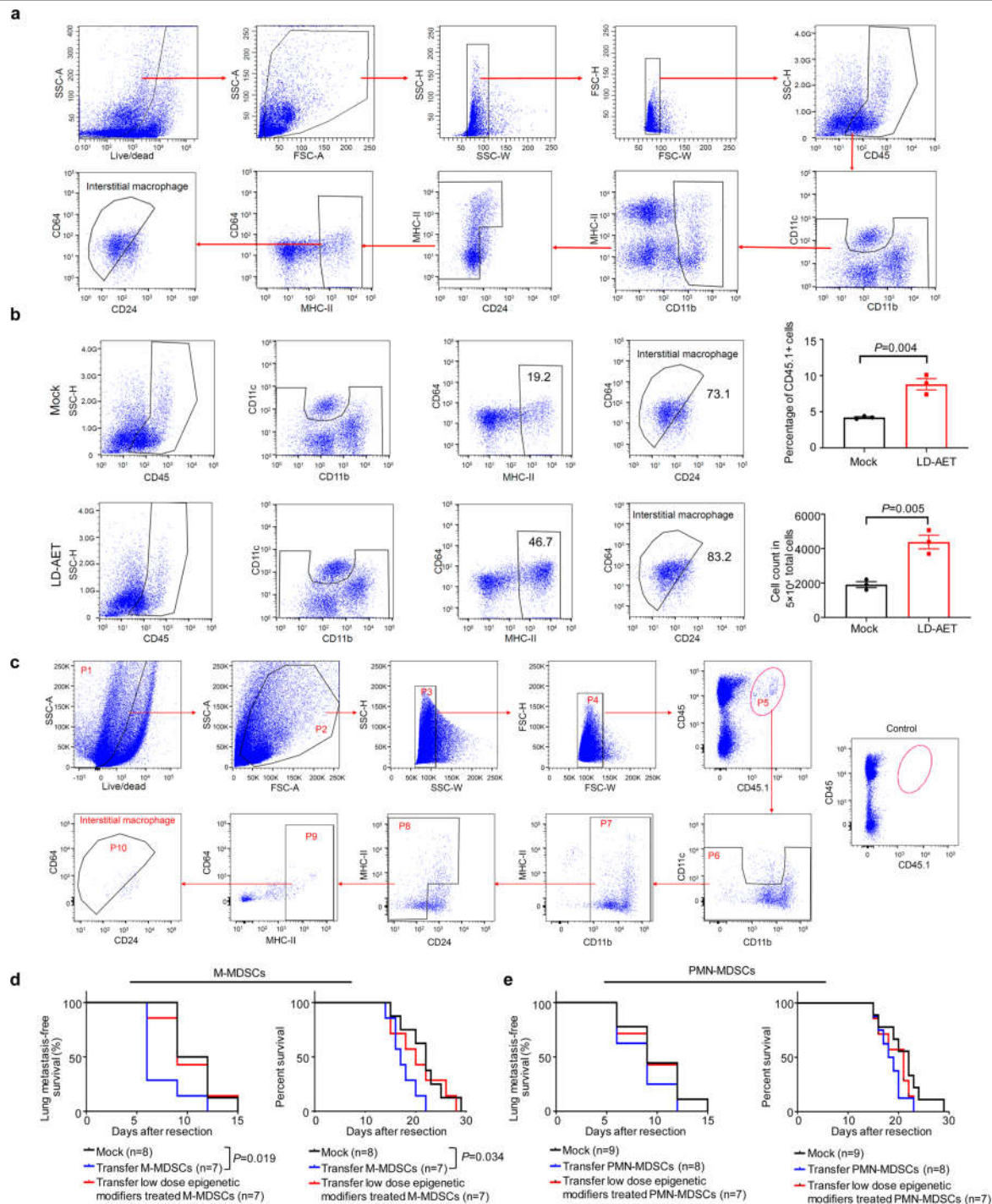


c



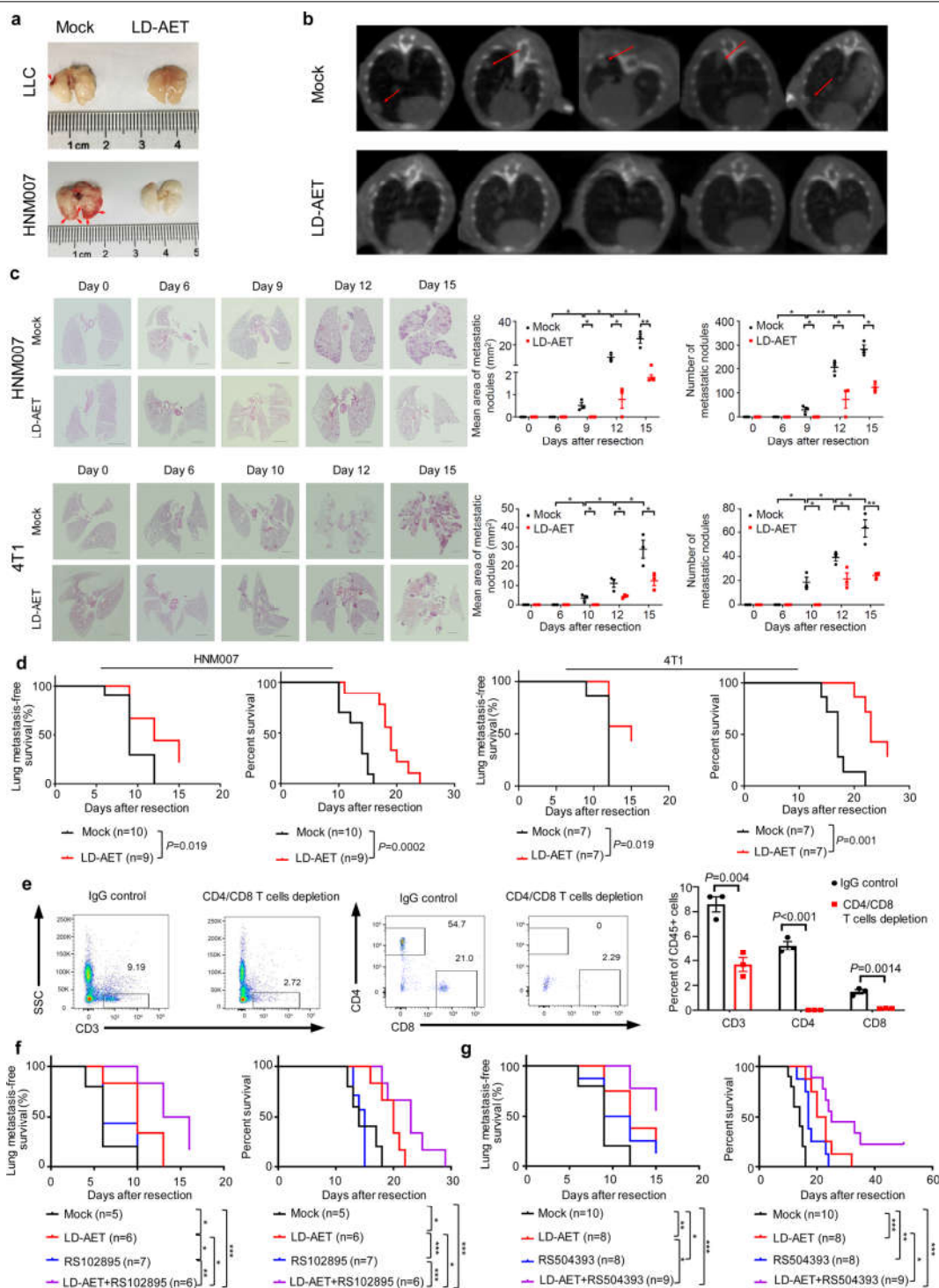
Extended Data Fig. 7 | Low-dose AET promotes the differentiation of monocytic MDSCs towards macrophages in the LLC model. **a**, DAVID analyses of the significantly downregulated and upregulated genes using the KEGG Gene Ontology in LLC mice, treated or untreated with low-dose AET ($n = 3$ biological replicates). Hypergeometric test (FDR-adjusted $P < 0.05$). **b**, **c**, The mRNA (**b**) and protein (**c**) levels of representative transcription factors were measured by quantitative PCR and western blot, respectively. In vitro,

splenic monocytic MDSCs from LLC mice were cultured for 3 days with tumour-conditioned medium. In vivo, monocytic MDSCs from both lungs of mock- (day 3) and low-dose-AET-treated LLC mice (day 3) were sorted for analysis. For gel source data, see Supplementary Fig. 1. Representative data were repeated at least three times with similar results. In **b**, a two-sample two-sided t -test was used, $n = 3$ biological replicates. All bars show mean \pm s.e.m.



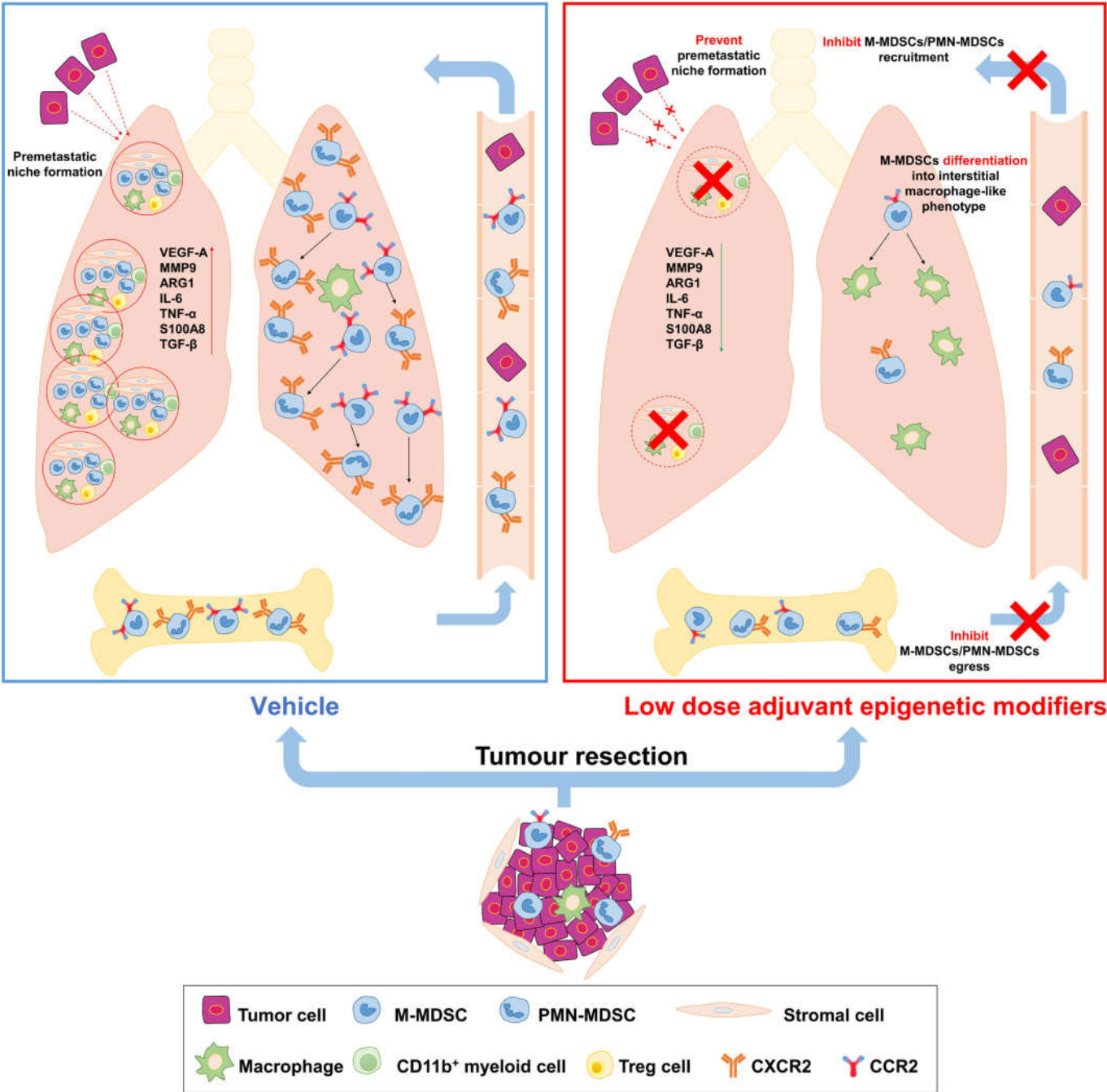
Extended Data Fig. 8 | Low-dose AET promotes the differentiation of monocytic MDSCs towards an interstitial macrophage-like population in the lung premetastatic microenvironment. **a**, Gating strategy used to identify and analyse lung interstitial macrophages in the lung premetastatic microenvironment by FACS. **b**, The effect of low-dose AET on lung interstitial macrophages from LLC mice. The percentage and cell counts of interstitial macrophages from both lungs in mock- and low-dose-AET-treated mice were analysed by FACS at day 3 after surgery ($n=3$ mice per group). Two-sample, two-sided t -test. All bars show mean \pm s.e.m. **c**, Gating strategy used to identify and analyse CD45.1⁺ lung interstitial macrophages from the lungs of recipient

CD45.2 mice after the transfusion of CD45.1⁺ monocytic MDSCs. **d**, Kaplan-Meier curves showing the disease-free survival and overall survival of *Ccr2*-knockout LLC mice after transfusion of wild-type monocytic MDSCs (5×10^6), low-dose-AET-treated (in vivo) wild-type monocytic MDSCs (5×10^6) or vehicle at day 1 and day 4, respectively. **e**, Kaplan-Meier curves showing the disease-free survival and overall survival of the *Ccr2*-knockout LLC mice after transfusion of wild-type polymorphonuclear MDSCs (5×10^6), low-dose-AET-treated (in vivo) wild-type polymorphonuclear MDSCs (5×10^6) or vehicle at day 1 and day 4, respectively. In **d**, **e**, two-sided log-rank tests were used.



Extended Data Fig. 9 | Low-dose AET inhibits pulmonary metastases and prolongs overall survival in mouse models. **a**, Representative photographs showing lungs treated with vehicle or low-dose AET in LLC (day 6) and HNM007 (day 10) mice. The red arrows indicate the metastases. **b**, Representative CBCT images of lung metastases on day 6 after resection in LLC mice treated with vehicle or low-dose AET. The red arrows indicate the metastases. **c**, Representative H&E-stained images of lung sections from HNM007 (top) and 4T1 (bottom) mice treated with low-dose AET or vehicle at different time points after surgery. Scale bars, 2 mm. Graph shows area and numbers of metastatic nodules. At each time point, three mice were killed for analysis. For each sample, sections from three levels were analysed. Two-sample, two-sided *t*-test. **d**, Kaplan-Meier curves showing the disease-free survival and overall survival of HNM007 and 4T1 mice, treated with low-dose AET (for the 4T1

model, 5-azacytidine 0.5 mg kg⁻¹ plus entinostat 2.5 mg kg⁻¹ d⁻¹) or vehicle after surgery. **e**, FACS showing the effect of T-cell-depleting antibodies on CD4⁺ and CD8⁺ T cells in the peripheral blood of LLC mice. *n* = 3 mice per group. Two-sample, two-sided *t*-test. **f**, Kaplan-Meier curves showing the disease-free survival and overall survival of LLC mice treated with vehicle, CCR2 antagonist (RS102895) (Sigma), low-dose AET and RS102895 plus low-dose AET after surgery. **g**, Kaplan-Meier curves showing the disease-free survival and overall survival of HNM007 mice treated with vehicle, CCR2 antagonist (RS504393) (Sigma), low-dose AET and RS504393 in combination with low-dose AET. In **d**, **f**, **g**, two-sided log-rank tests were used. Representative data in **a**, **b** were repeated at least three times with similar results. All bars show mean \pm s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.



Extended Data Fig. 10 | Representation of the effect of low-dose adjuvant epigenetic modifiers on lung metastases in the tumour pulmonary-metastasis models. Graphic model showing the inhibition of pulmonary metastases by low-dose AET via its effect on MDSCs. First, low-dose AET can inhibit the trafficking of monocytic and polymorphonuclear MDSCs from the bone marrow to the premetastatic microenvironment by downregulating the

expression of CCR2 and CXCR2, respectively. Second, even if MDSCs migrate to the lung, low-dose AET can skew the differentiation of monocytic MDSCs towards an interstitial macrophage-like phenotype in the lung premetastatic microenvironment. Therefore, low-dose AET can disrupt the lung premetastatic microenvironment, ultimately inhibiting pulmonary metastases.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

- 1.FACSARIA II SORP cell sorter, LSRII, FACS Diva(BD Biosciences); MoFlo XDP Cell Sorter (BECKMAN COULTER)
- 2.Cone beam computed tomography (CBCT) guided systems(Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University)
- 3.Nikon Eclipse NiE microscope (Nikon Instruments Inc.)
- 4.NanoDrop One(Thermo Scientific)
- 5.Bio-Rad CFX96 Real-Time System and Bio-Rad CFX Manager v2.1 (BIO-RAD)
- 6.Bio-Rad TC20 Automated Cell Counter(BIO-RAD)

Data analysis

GraphPad Prism version 7.0, FlowJo software version 10.0.7, Photoshop(Adobe),Aperio Imagescope software version 12.4, R version 3.6.1, EZ-C1 software version 3.90, Broad Institute GSEA 3.0, DAVID version 6.8, Summit 5.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated are included in the published Letter and in its Supplementary information. Gene-expression data that support the findings of this study have been deposited in the Gene Expression Omnibus under accession numbers GSE124539. All data are also available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. Sample sizes were selected empirically from previous experimental experience with similar assays, and/or from sizes generally employed in the field.
Data exclusions	No exclusion of data was made.
Replication	Experiments were performed at least 3 times and/or with sufficient cells/animals per group to demonstrate statistical significance. All attempts at replication were successful.
Randomization	This is applicable to the in vivo animal experiments: Mice were divided randomly into cages before tumor resection and treated with corresponding drugs postresection.
Blinding	The performer of the animal experiment was blinded to the randomization process. Pathological examination was done in a blinded fashion.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Antibodies were used for flow cytometry, immunoblotting, immunofluorescence and ELISA.

- Flow cytometry was conducted using fluorochrome-conjugated antibodies.
- For staining of mouse cells we used: anti-CD45(Biolegend, 103116),anti-CD11b(BD biosciences, 563168),anti-Gr-1(BD biosciences, 553127),anti-CD3(BD biosciences, 552774),anti-CD4(BD biosciences, 562891),anti-CD8a(Biolegend, 100711),anti-CD25(BD biosciences, 552880),anti-Foxp3(BD biosciences, 560414),anti-CD45.1(Biolegend, 110708),anti-CD192(Biolegend, 150610),anti-Ly6C(Biolegend, 128008),anti-Ly6G(Biolegend, 127606),anti-I-A/I-E(BD biosciences, 562564),anti-CD11c(BD biosciences, 550261),anti-CD24(BD biosciences, 564237),anti-CD64(BD biosciences, 740622),anti-CD16/32(Biolegend, 101320),anti-F4/80(Biolegend, 123145),anti-CXCR1(BD biosciences, 566383),anti-CXCR2(Biolegend, 149609),anti-Ki67(Biolegend, 652409). All antibodies were used at 1:100 dilution for flow cytometry staining.
- Immunoblot: anti-MMP9(abcam, ab38898),anti-TGF-beta(abcam, ab179695),anti-ARG1(abcam, ab91279),anti-VEGF-A(abcam, ab214424),anti-S100A8(abcam, ab92331),anti-TNF-alpha(abcam, ab183218),anti-IL-6(Cell Signaling Technology, 12912),anti-EGFR1(Abcam, ab133695),anti-EGFR2(Abcam, ab108399),anti-PPAR-gamma(Cell Signaling Technology, 2435),anti-Maf-b(Santa Cruz Biotechnology, sc-376387),anti-p50(Cell Signaling Technology, 55764),anti-p52(Cell Signaling Technology, 4882),anti-Rel-B(Cell Signaling Technology, 55764),anti-p65(Cell Signaling Technology, 55764),anti-beta-actin(Sigma, A5441),anti-DNMT1(Cell Signaling Technology, 5032),and anti-rabbit IgG, HRP-linked antibody and anti-mouse IgG, HRP-linked antibody(Cell Signaling Technology, 7074, 7076). All primary antibodies were used at 1:1000 dilution for immunoblotting, except anti-Maf-b was used at 1:100 and anti-beta-actin was used at 1:10000. Secondary antibodies were used at a 1:2000 dilution.
- Immunofluorescence analysis: Anti-Gr-1 antibody (Biolegend, 108401, 1:500),anti-GFP antibody(Abcam, ab183734, 1:200), anti-CD4 antibody (Abcam, ab183685, 1:200), anti-CD8a antibody (ebioscience, 14-0808-80, 1:100),and goat anti-rabbit and anti-rat secondary antibodies(Invitrogen, A27034, A18876).
- IFN-gamma ELISA kits: Mouse IFN-gamma (DY485) from R&D.
- For in vivo CD4/CD8 depletion we used: anti-mouse CD4 antibody(BE0003-1) , anti-mouse CD8a antibody(BE0118) and isotype purchased from BioXcell.

Validation

All antibodies are commercially available and validated by the manufactures for the applications and species used in this study. See manufactures websites for validation statements (www.abcam.com; www.cellsignal.com; <https://www.bdbiosciences.com/en-us>; <https://www.legends.com>; <https://www.bxcell.com>; <https://www.thermofisher.com>; <https://www.sigmaldrich.com>; <https://www.rndsystems.com>)

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Lewis lung carcinoma cells (LLC1) and 4T1 cells were obtained from ATCC.HNM007, a p53-null murine esophageal squamous cell carcinoma cell line transformed by HRASG12V, was kindly provided by Prof. S Singhal (University of Pennsylvania).
Authentication	No method authentication was performed for the mouse cell lines used.
Mycoplasma contamination	Cell lines were routinely tested for mycoplasma and immediately tested upon suspicion. None of the lines used in the reported experiments tested positive.
Commonly misidentified lines (See ICLAC register)	None were misidentified.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	6-8 week-old female C57BL/6 mice and BALB/c mice were purchased from Charles River Laboratories (Wilmington, MA). 6-8 week-old female NOD/SCID/g-chain knockout (NSG) mice, bred and housed at the Johns Hopkins Animal Care Facility, were used. 6-8 week-old female B6.129S4-Ccr2tm1fc/J mice were kindly provided by Prof. Sharon A. McGrath-Morrow (Johns Hopkins University). 6-8 week-old female mice congenic in murine CD45 at the Ly5 locus (B.6SJL-Ptprca Pepcb/BoyJ Ly5.1) were obtained from Jackson Laboratory.
Wild animals	No wild animals were used in the study.
Field-collected samples	No field-collected samples were used in the study.
Ethics oversight	The Guide for the Care and Use of Laboratory Animals published by the National Institutes of Health (NIH publication 96-01, revised 1996) were followed, and all protocols were approved by the Johns Hopkins University Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	13 patients(4 female, 9 male) with stage I non-small cell lung cancer and an Eastern Cooperative Oncology Group (ECOG) performance status of 0-1 after curative surgery.
Recruitment	Patients with stage I NSCLC who presented at John Hopkins for medical care and fulfilled the eligibility criteria as per approved protocol were offered participation in the study. We are unaware of any potential self-selection bias or other biases present.
Ethics oversight	Human biological samples were sourced ethically and their research use was in accord with the terms of the informed consent provided under the institutional protocols approved either by Johns Hopkins Hospital Ethics Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	NCT01207726
Study protocol	Study protocol is accessible on Clinical Trials website. More information is available from the corresponding author upon reasonable request.
Data collection	The J1037 (NCT01207726) study was a randomized phase II study that compared the low dose adjuvant epigenetic therapy (Aza plus entinostat) with standard of care (observation) in patients with stage I (T1-2aN0) NSCLC after primary tumour resection in Johns Hopkins Hospital and Anne Arundel Medical Center. The patients from the adjuvant epigenetic treatment group received the combination of azacitidine at 40 mg/m ² on days 1-5 and 8-10 with entinostat at a 7 mg fixed dose on days 3 and 10 of each 28 days cycle. The primary end point was the effect of Aza plus entinostat on the hazard of 3 years progression-free survival in patients with resected stage I non-small cell lung cancer. Finally, thirteen patients were enrolled in the trial. Due to the difficulty in enrolling patients, the trial was prematurely terminated on May 1, 2015.

Outcomes

Primary Outcome Measures :

1. Disease-free Survival (DFS) [Time Frame: 3 years]

The DFS hazard rate and 95% confidence interval will be reported. At this time, event time distributions for disease-free survival in the two arms will be estimated with the method of Kaplan and Meier and compared using a stratified Cox-proportional hazards model (stratified for stage IA vs IB) with a two-sided alpha of 10%.

Secondary Outcome Measures :

1. Factors That Predict Clinical Outcome in Patients Treated With Combination Epigenetic Therapy in Terms of Epigenomic Data Generated From the Illumina Platform [Time Frame: Up to 2 years]

The study was terminated early due to poor accrual since the requirement of clinic administration of the 5AZA daily and post-operative patients not wanting 6 months of treatment. For this reason, 13 pts were enrolled and data was not analyzed, for which we are unable to make any conclusions or report results.

2. Median Disease-free Survival [Time Frame: Up to 5 years]

Determined by the method determined by Kaplan and Meier. Estimated with 95% confidence intervals. Cox proportional hazard modeling will be used for multivariate analysis.

3. Number of Relapses and Deaths Per Total Time of Follow-up Comparing Patients With N2 Lymph Nodes in Terms of Methylated and Unmethylated [Time Frame: Up to 5 years]

Kaplan Meier curves will be used.

4. Overall Survival [Time Frame: Up to 5 years]

Determined by the method determined by Kaplan and Meier. Estimated with 95% confidence intervals. Cox proportional hazard modeling will be used for multivariate analysis.

5. Presence of Methylation Patterns [Time Frame: Up to 2 years]

McNemar's test will be used to compare the change in methylation after treatment in sputum.

6. Toxicities Graded According to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) v4.0 [Time Frame: Up to 5 years]

Simple descriptive statistics will be utilized to display the data.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	See methods
Instrument	Cell were recorded on an LSR II Flow Cytometer (BD Biosciences), or on a MoFlo XDP Cell Sorter (BECKMAN COULTER)
Software	Data were analyzed using FlowJo v10.0.0 software, or Summit 5.4 software.
Cell population abundance	For sorting experiments, the cell were subsequently analyzed by FACS and >98% purity was achieved.
Gating strategy	Following routine gating strategy.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

DNA-PKcs has KU-dependent function in rRNA processing and haematopoiesis

<https://doi.org/10.1038/s41586-020-2041-2>

Received: 16 March 2019

Accepted: 28 January 2020

Published online: 26 February 2020

 Check for updates

Zhengping Shao^{1,2,11}, Ryan A. Flynn^{3,11}, Jennifer L. Crowe^{1,4,11}, Yimeng Zhu^{1,2,11}, Jialiang Liang⁵, Wenxia Jiang^{1,2}, Fardin Aryan⁵, Patrick Aoude⁵, Carolyn R. Bertozzi^{3,6}, Verna M. Estes^{1,2}, Brian J. Lee^{1,2}, Govind Bhagat^{2,7,8,9}, Shan Zha^{1,2,7,8,9}✉ & Eliezer Calo^{5,10}✉

The DNA-dependent protein kinase (DNA-PK), which comprises the KU heterodimer and a catalytic subunit (DNA-PKcs), is a classical non-homologous end-joining (cNHEJ) factor¹. KU binds to DNA ends, initiates cNHEJ, and recruits and activates DNA-PKcs. KU also binds to RNA, but the relevance of this interaction in mammals is unclear. Here we use mouse models to show that DNA-PK has an unexpected role in the biogenesis of ribosomal RNA (rRNA) and in haematopoiesis. The expression of kinase-dead DNA-PKcs abrogates cNHEJ². However, most mice that both expressed kinase-dead DNA-PKcs and lacked the tumour suppressor TP53 developed myeloid disease, whereas all other previously characterized mice deficient in both cNHEJ and TP53 expression succumbed to pro-B cell lymphoma³. DNA-PK autophosphorylates DNA-PKcs, which is its best characterized substrate. Blocking the phosphorylation of DNA-PKcs at the T2609 cluster, but not the S2056 cluster, led to KU-dependent defects in 18S rRNA processing, compromised global protein synthesis in haematopoietic cells and caused bone marrow failure in mice. KU drives the assembly of DNA-PKcs on a wide range of cellular RNAs, including the U3 small nucleolar RNA, which is essential for processing of 18S rRNA⁴. U3 activates purified DNA-PK and triggers phosphorylation of DNA-PKcs at T2609. DNA-PK, but not other cNHEJ factors, resides in nucleoli in an rRNA-dependent manner and is co-purified with the small subunit processome. Together our data show that DNA-PK has RNA-dependent, cNHEJ-independent functions during ribosome biogenesis that require the kinase activity of DNA-PKcs and its phosphorylation at the T2609 cluster.

The cNHEJ pathway directly ligates DNA double-strand breaks (DSBs) and is required for V(D)J recombination in developing lymphocytes⁵. The KU70–KU86 (Ku80 in mice) heterodimer (KU) binds to DNA ends, promotes ligation by the DNALIG4–XRCC4–XLF complex¹, and recruits and activates DNA-PKcs, which activates the Artemis endonuclease for end-processing¹. DNA-PKcs is required for end-processing, but not blunt-end ligation^{6–8}. Productive V(D)J recombination requires both end-processing and end-ligation⁵. Correspondingly, DNA-PKcs-null mice lack mature lymphocytes, but are otherwise normal^{6–8}.

To understand how DNA-PK activity regulates DNA-PKcs, the best characterized substrate of DNA-PK, we generated *DNA-PKcs*^{KD/KD} mice² (also known as *Prkdc*^{KD/KD}). *DNA-PKcs*^{KD/KD} mice died embryonically with TP53-dependent neuronal apoptosis², similar to end-ligation-deficient *Xrcc4*^{−/−} or *Lig4*^{−/−} mice^{9,10}, indicating that, once recruited to DNA ends, DNA-PKcs requires its kinase activity to license end-ligation. In a TP53-deficient background, cNHEJ-deficient mice (such as *Xrcc4*^{−/−} mice)

succumb to pro-B cell lymphomas by the age of 100 days^{3,9,11} (Fig. 1a–c). Unexpectedly, most *DNA-PKcs*^{KD/KD} *TP53*^{−/−} mice (*TP53* is also known as *Trp53*) died by 40 days without lymphomas (fewer than 25% survived for more than 80 days and developed pro-B cell lymphomas; Fig. 1a–c, Extended Data Fig. 1a). Pro-B cell lymphomas from *Xrcc4*^{−/−} *TP53*^{−/−} mice carry an alternative end-joining (alt-EJ)-mediated immunoglobulin *Igh*/*Myc* translocation and co-amplification³, leading to overexpression of MYC. Alt-EJ is not affected in *DNA-PKcs*^{KD/KD} *TP53*^{−/−} B cells¹², so lack of translocation cannot explain the low incidence of lymphomas in *DNA-PKcs*^{KD/KD} *TP53*^{−/−} mice. Moreover, the early death of *DNA-PKcs*^{KD/KD} *TP53*^{−/−} mice is not due to incomplete rescue of the lethal *DNA-PKcs*^{KD/KD} phenotype by lack of TP53, as most *DNA-PKcs*^{KD/KD} *TP53*^{−/−} mice survived to adulthood before succumbing to thymic lymphomas and/or sarcomas due to *TP53* deficiency¹³ (Fig. 1a–c). Bone marrow and spleen samples from young *DNA-PKcs*^{KD/KD} *TP53*^{−/−} mice revealed a myeloid expansion, accumulation of hypolobated and hyperchromatic

¹Institute for Cancer Genetics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ²Herbert Irving Comprehensive Cancer Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ³Department of Chemistry, Stanford University, Stanford, CA, USA. ⁴Graduate Program of Pathobiology and Molecular Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁶Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. ⁷Department of Pediatrics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ⁸Department of Pathology and Cell Biology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ⁹Department of Immunology and Microbiology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ¹⁰Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. ¹¹These authors contributed equally: Zhengping Shao, Ryan A. Flynn, Jennifer L. Crowe, Yimeng Zhu. ✉e-mail: sz2296@columbia.edu; calo@mit.edu

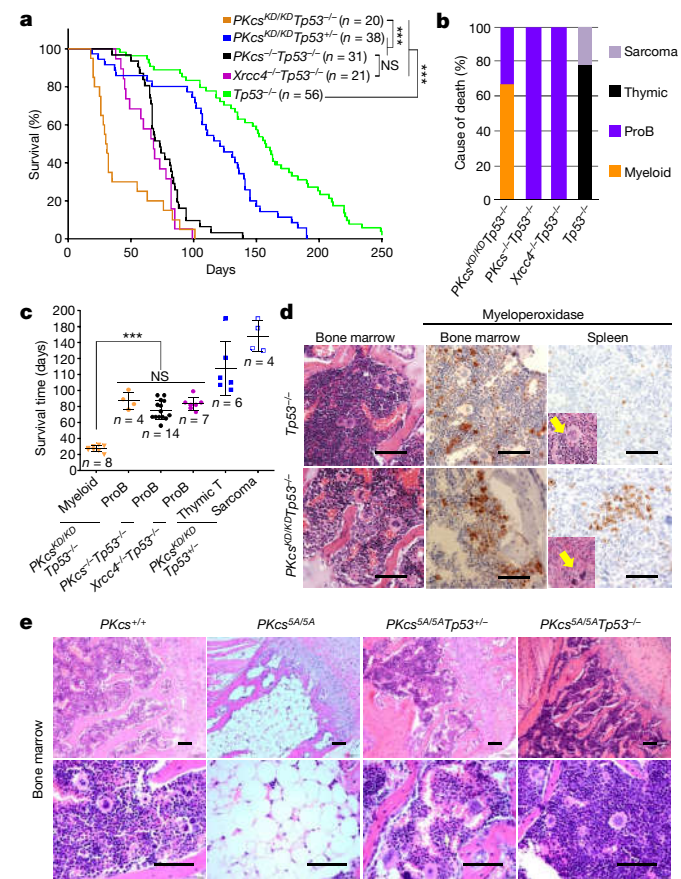


Fig. 1 | Kinase dead or phosphorylation deficient DNA-PKcs causes haematopoietic defects independent of cNHEJ. **a**, Kaplan–Meier survival curve of *DNA-PKcs^{KD/KD}TP53^{-/-}* and control (*DNA-PKcs^{KD/KD}TP53^{+/+}*, *DNA-PKcs^{-/-}TP53^{-/-}*, *Xrcc4^{-/-}TP53^{-/-}* and *TP53^{-/-}*) mice. Log-rank (Mantel–Cox) test; ****P* < 0.001, NS (not significant) *P* > 0.05. **b**, Cause of death for tumour cohorts by genotype. **c**, The life expectancy for mice of different genotypes plotted by tumour type. Pro-B cell lymphomas in *DNA-PKcs^{KD/KD}TP53^{-/-}* mice have similar latency as in *DNA-PKcs^{-/-}TP53^{-/-}* and *Xrcc4^{-/-}TP53^{-/-}* mice (two-sided unpaired Student's *t*-test; ****P* < 0.001, NS *P* > 0.05). Data shown as mean ± s.e.m. **d**, Representative histological analyses of bone marrow and spleen from 3-week-old *DNA-PKcs^{KD/KD}TP53^{-/-}* mice. The samples were stained for the myeloid marker myeloperoxidase before being counterstained with haematoxylin (for nuclei). Arrows indicate megakaryocytes. Scale bars, 100 μm. **e**, Representative histology analyses of bone marrow from 2-week-old *DNA-PKcs^{+/+}* and *DNA-PKcs^{SA/SA}* mice with or without *TP53* deficiency. Scale bars, 100 μm. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.

megakaryocytes, and occasionally c-KIT-positive myeloid blasts (Fig. 1d, Extended Data Fig. 1b, c), characteristic of myelodysplastic syndrome (MDS). MDS does not manifest in other cNHEJ- and TP53-deficient mice, suggesting that kinase-dead DNA-PKcs might have a role in haematopoiesis beyond cNHEJ.

DNA-PKcs mutations affect haematopoiesis

To investigate whether loss of autophosphorylation of DNA-PKcs explains the haematopoietic defects, we generated mice with alanine substitutions to either of the two autophosphorylation clusters^{14,15}—the T2609 cluster (*DNA-PKcs^{SA}*) or the S2056 cluster (*DNA-PKcs^{PQR}*)¹⁶ (Extended Data Fig. 2a–d). *DNA-PKcs^{PQR/PQR}* mice were healthy¹⁶ with normal red blood cell (RBC) and platelet counts (Extended Data Fig. 2e, f). By contrast, alanine substitutions at all five (*DNA-PKcs^{SA/SA}*) or three out of five (*DNA-PKcs^{3A/3A}*)¹⁷ threonines in the T2609 cluster caused TP53-dependent bone marrow failure (Fig. 1e) and death by 4 weeks of age

(Fig. 2a) with severe anaemia, pale organs and thrombocytopenia (Fig. 2b, Extended Data Fig. 2g–i). Moreover, haematopoietic stem and progenitor cells (HSPCs; Lin[−]SCA1⁺c-KIT⁺) and their downstream progenitors (Lin[−]SCA1[−]c-KIT⁺) decreased by around 1,000-fold in *DNA-PKcs^{SA/SA}* mice (Fig. 2c, d, Extended Data Fig. 3a, b). Fetal liver HSPC counts from *DNA-PKcs^{SA/SA}* and *DNA-PKcs^{KD/KD}* mice also decreased by half (Extended Data Fig. 3c, d), suggesting that these mice had an early haematopoietic defect. Thus, both the kinase activity of DNA-PKcs and autophosphorylation at the T2609 cluster are necessary for haematopoiesis in mice expressing DNA-PKcs.

Effect of DNA-PK mutations on DNA repair

Next, we investigated whether the anaemia in *DNA-PKcs^{SA/SA}* mice was triggered by defects in known functions of DNA-PKcs. *DNA-PKcs^{SA/SA}TP53^{-/-}* mice were viable¹⁷ (Fig. 2a) and had wild-type levels of myeloid cells, neutrophils, erythrocytes and platelets, and notably had a substantial number of mature B and T lymphocytes (Extended Data Fig. 4a–e). By contrast, *DNA-PKcs^{-/-}TP53^{-/-}* and *DNA-PKcs^{KD/KD}TP53^{-/-}* mice² lacked mature B and T lymphocytes (Extended Data Figs. 1a, 4a, b), suggesting that DNA-PKcs^{SA} supports lymphocyte development. Accordingly, V(D)J recombination occurred efficiently in *DNA-PKcs^{SA/SA}* (Extended Data Fig. 4f–h) and *DNA-PKcs^{3A/3A}* B cells¹⁸. T cell development was also unaffected in *DNA-PKcs^{SA/SA}TP53^{-/-}* mice (Extended Data Fig. 4i). Moreover, deletion of KU, which abrogates DNA-PK assembly and cNHEJ, rescued lethality (Fig. 2a) and restored HSPCs, RBCs and neutrophils, but not lymphocytes, in *DNA-PKcs^{SA/SA}* mice (Fig. 2b–d, Extended Data Fig. 4d, e), suggesting that DNA-PKcs has cNHEJ-independent functions during haematopoiesis that require phosphorylation of the T2609 cluster. Accordingly, *DNA-PKcs^{SA/SA}* mouse embryonic stem cells (ES cells) and fibroblasts (MEFs) were not markedly hypersensitive to DNA-damaging agents such as mitomycin C (MMC) or ionizing radiation (IR)^{6,19} (Extended Data Fig. 5a–e). *DNA-PKcs^{SA/SA}* MEFs were more resistant to MMC than MEFs from mice deficient in the Fanconi anaemia (FA) pathway (*Fancd2^{-/-}*; Extended Data Fig. 5b). Unlike humans with *FANCD2* deficiency, *Fancd2^{-/-}* mice do not develop spontaneous lethal anaemia¹⁹. These findings suggest that defects in the FA pathway cannot explain the lethal bone marrow failure of *DNA-PKcs^{SA/SA}* mice. *DNA-PKcs^{SA/SA}* MEFs proliferated well, with normal S-phase frequency (Extended Data Fig. 5f, g), in contrast to the marked proliferation defects of cNHEJ-deficient MEFs^{6,10}. On the other hand, *DNA-PKcs^{SA/SA}* ES cells had reduced S-phase and formed smaller colonies (Extended Data Fig. 6a–c), whereas cNHEJ-deficient ES cells proliferate well^{6,10}. CRISPR-mediated deletion of *Ku80* rescued the colony-formation defects of *DNA-PKcs^{SA/SA}* ES cells (Extended Data Fig. 6a, b). Telomere stability^{20,21} was comparable between *DNA-PKcs^{SA/SA}* and *DNA-PKcs^{-/-}* cells (Extended Data Fig. 6d–f). These findings and the fact that haematopoiesis is unaffected in age-matched *DNA-PKcs^{-/-}* mice (Fig. 2b, Extended Data Fig. 2g), suggest that neither telomere dysfunction nor cNHEJ deficiency alone can explain the lethal anaemia in *DNA-PKcs^{SA/SA}* mice.

DNA-PK mutations compromise translation

In addition to DNA, purified KU also binds to RNA²². Yeast Ku binds to the telomerase RNA component (TERC)^{23,24}, but the RNA partners of mammalian KU and the relevance of KU–RNA interactions in mammalian cells remain unclear. Without DNA damage, KU and DNA-PKcs, but not other cNHEJ factors, reside in the nucleolus in a detergent-resistant and Pol I transcription-sensitive manner²⁵ (Extended Data Fig. 8a–c), suggesting that DNA-PK has an RNA-dependent role in the nucleolus. Defects in ribosome biogenesis can trigger TP53-dependent haematopoietic failure²⁶, and protein synthesis is tightly regulated in HSPCs²⁷. Defects in protein synthesis, as exemplified in Diamond–Blackfan anaemia (DBA)²⁸, cause

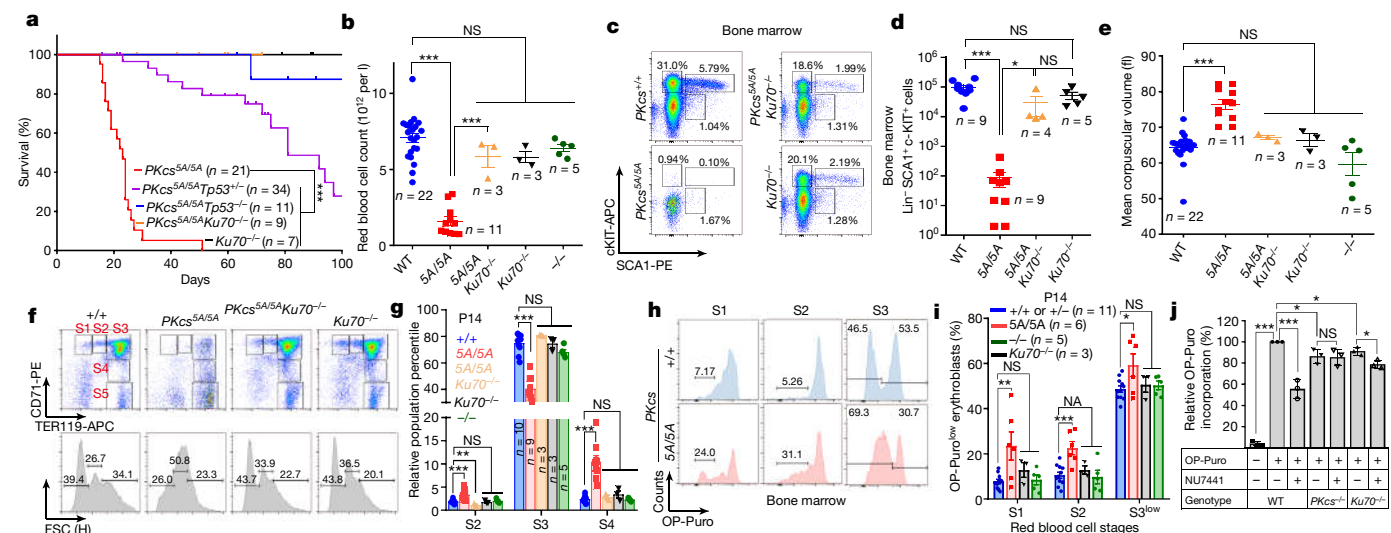


Fig. 2 | Mutations in DNA-PKcs cause KU-dependent haematopoietic failure and translation defects. **a**, Kaplan–Meier survival curve of *DNA-PKcs*^{SA/SA}, *DNA-PKcs*^{SA/SA}*Tp53*^{-/-}, *DNA-PKcs*^{SA/SA}*Tp53*^{-/-}, *DNA-PKcs*^{SA/SA}*Ku70*^{-/-} and *Ku70*^{-/-} mice. **b**, *Ku70* deficiency rescued the peripheral RBC counts of 2-week-old *DNA-PKcs*^{SA/SA}*Ku70*^{-/-} mice. **c**, **d**, Representative flow cytometry analyses (**c**) and the absolute number of haematopoietic stem and progenitor cells (Lin⁻SCA1⁺c-KIT⁺) (**d**) from 2-week-old *DNA-PKcs*^{SA/SA} and *DNA-PKcs*^{SA/SA}*Ku70*^{-/-} mice. **e**, MCV of RBCs from 2-week-old *DNA-PKcs*^{SA/SA} and *DNA-PKcs*^{SA/SA}*Ku70*^{-/-} mice. **f**, Representative flow cytometry analyses of bone marrow RBCs from 2-week-old *DNA-PKcs*^{SA/SA} and *DNA-PKcs*^{SA/SA}*Ku70*^{-/-} mice. CD71 and TER119 staining determines the stage of RBC differentiation. S1, CD71⁺TER119^{low}; S2, CD71⁺TER119^{mid}; S3, CD71⁺TER119^{high}; S4, CD71^{mid}TER119^{high}; S5, CD71⁺TER119^{high}. Forward scatter of TER119⁺ RBCs is shown (bottom). **g**, Relative frequencies of S2, S3 and S4 among all immature erythroblasts

TP53-dependent macrocytic anaemia with increased mean corpuscular volume (MCV, RBC size). *DNA-PKcs*^{SA/SA} mice but not *DNA-PKcs*^{-/-} mice displayed TP53-dependent macrocytic anaemia (30% increase in MCV; Fig. 2b, e, Extended Data Figs. 4c, 6g, 7a). In *DNA-PKcs*^{SA/SA} mice the percentage of CD71⁺TER119^{high} (S3) erythroblasts was lower in fetal liver at embryonic day 14.5 (E14.5) and in bone marrow at postnatal day 14 (P14) than in wild-type mice (Fig. 2f, g, Extended Data Fig. 7b, c). Notably, CD71⁺ erythroblasts (S1–S3) have the highest translation levels among all the haematopoietic cell-types examined (Fig. 3f, Extended Data Figs. 3e, 7d), as measured by the incorporation of O-propargyl-puromycin (OP-Puro) into nascent proteins²⁷. An OP-Puro^{low} population accumulated throughout erythroblast development in the P14 bone marrow (Fig. 2h, i) and fetal liver (Extended Data Fig. 7e, f) of *DNA-PKcs*^{SA/SA} mice, suggesting that there were global translation defects. *DNA-PKcs*^{SA/SA}*Tp53*^{-/-} erythroblasts also displayed consistent translation defects (Extended Data Fig. 7g), suggesting that DNA-PKcs has a TP53-independent role in protein synthesis. Likewise, TP53-deficient, v-ABL kinase-transformed pro-B cell lines²⁹ derived from *DNA-PKcs*^{KD/KD} or *DNA-PKcs*^{SA/SA} mice showed reduced translation (Extended Data Fig. 7h). Chemical inhibition of DNA-PKcs (with NU7441), but not of the related ataxia telangiectasia-mutated (ATM) kinase (with KU55933), reduced global translation (Fig. 2j, Extended Data Fig. 7i). *DNA-PKcs*^{-/-} and *Ku70*^{-/-} (also known as *Xrcc6*) B cells had moderate but consistent translation defects that were insensitive to inhibition of DNA-PK (Fig. 2j). Erythrocyte counts and translation of S3 erythroblasts in young *DNA-PKcs*^{-/-} or *Ku70*^{-/-} mice were comparable to those in wild-type mice, and significantly ($P < 0.05$) higher than in *DNA-PKcs*^{SA/SA} mice (Fig. 2b, f–i, Extended Data Fig. 6g, h). These results reveal a critical role for DNA-PK kinase and phosphorylation of the T2609 cluster during translation in HSPCs in the presence of KU and DNA-PKcs.

(S1–S4). The S5 percentage is much higher in *DNA-PKcs*^{SA/SA} mice (Extended Data Fig. 7c). **h**, Measurements of translation in S1, S2 and S3 erythroblasts from P14 *DNA-PKcs*^{+/+} and *DNA-PKcs*^{SA/SA} mice bone marrow. **i**, Quantification of the frequency of OP-Puro^{low} among S1, S2 and S3 erythroblasts by OP-Puro labeling of bone marrow cells for 1 h. **j**, Inhibition of DNA-PKcs with NU7441 leads to a DNA-PKcs- and Ku-dependent reduction in global translation. Graphs show fold change in OP-Puro fluorescence (mean \pm s.d.) normalized to *DNA-PKcs*^{+/+} cells (set to 100%). Wild-type with or without NU7441 (44% reduction with NU7441); *DNA-PKcs*^{-/-} with or without NU7441 (1% reduction with NU7441), and *Ku70*^{-/-} with or without NU7441 (10% reduction with NU7441). **b**, **d**, **e**, **g**, **i**, Mean \pm s.e.m. **a**, Two-sided log-rank (Mantel–Cox) test, *** $P < 0.001$. **b**, **d**, **e**, **g**, **i**, Two-sided paired Student's *t*-test; all other panels, two-sided unpaired Student's *t*-test, *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; NS, $P > 0.05$. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.

DNA-PK binds to RNA

The biological importance of KU in the nucleolus is unknown. We hypothesized that defects in ribosome biogenesis might underlie the translation defects in *DNA-PKcs*^{KD/KD} and *DNA-PKcs*^{SA/SA} HSPCs. Indeed, blocking the activity of RNA polymerase I (Pol I) with actinomycin D (ActD) depleted KU and DNA-PKcs from the nucleolus in human and mouse cells (Extended Data Fig. 8a–c), suggesting that DNA-PKcs and KU reside in the nucleolus in an rRNA-dependent manner, potentially as part of pre-rRNA ribonucleoprotein complexes. Thus, we isolated the small subunit (SSU) processome via the U3 small nucleolar RNA (snoRNA; herein U3). U3 coordinates splicing of the 5'-external transcribed spacer (5'-ETS) and thus maturation of the 40S ribosomal subunit³⁰. We identified U3 binding proteins using comprehensive identification of RNA-binding proteins by mass spectrometry (ChIRP-MS)³¹ in two different cell types (Fig. 3a, b, Extended Data Fig. 8d, e). ChIRP-MS recovered all known and conserved components of the eukaryotic SSU processome^{4,32} and the DNA-PK holoenzyme, but not other cNHEJ factors (Fig. 3a, b, Supplementary Table 1). Comparative analyses of the U3 ChIRP-MS and KU86 immunoprecipitation (IP)-MS³³ confirmed that KU86 associates with SSU components (Extended Data Fig. 9a). If the assembly of DNA-PK^{KD} or DNA-PK^{SA} affects rRNA processing, unprocessed rRNA intermediates might accumulate in DNA-PK mutant cells. Northern blots from *DNA-PKcs*^{KD/KD}, *DNA-PKcs*^{SA/SA} and *DNA-PKcs*^{SA/SA} cells revealed partial accumulation of the 21S and 12S pre-rRNA precursors of the 18S and 5.8S rRNAs, respectively (Fig. 3c, d, Extended Data Fig. 9b). Given that *DNA-PKcs* mutant cells were viable, we expected the rRNA processing defects to be less severe than and/or different from those resulting from SSU ablation. Notably, deletion of *Ku70* along with mutations in *DNA-PKcs* rescued the rRNA processing defects (Fig. 3c, d, Extended Data Fig. 9b). The 21S pre-rRNA intermediate also

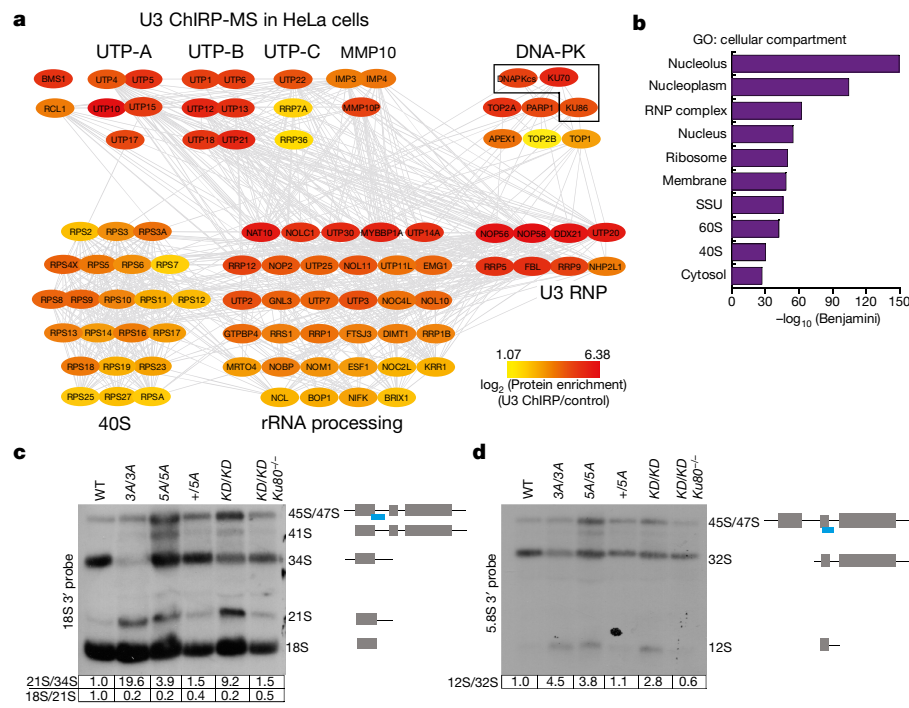


Fig. 3 | DNA-PK, but not other cNHEJ factors, co-purifies with the U3 snoRNA and regulates rRNA processing. **a**, A protein–protein interaction (grey lines) network of U3 ChIRP-MS from HeLa cells of nucleolar and ribosomal biogenesis factors. Colours map to the enrichment $\log_2[\text{ChIRP}/\text{control}]$ value of each protein. Known complexes are physically grouped and labelled. Complete hit list detailed in Supplementary Table 1. **b**, Cellular compartment term analysis of the U3 ChIRP-MS from HeLa cells using the DAVID tool. False discovery rate

(FDR) was estimated (Benjamini corrected P) by comparing the 483 enriched proteins to the human proteome; the top ten most enriched terms are shown. **c**, **d**, Northern blot analyses of 18S and 5.8S rRNA maturation in v-ABL kinase transformed B cells from noted genotypes. The probe covers the 3' edge of 18S rRNA and 5.8S rRNA, respectively (blue line). The normalized relative intensities of 21S/34S, 18S/21S and 12S/32S are marked below the gels. $n = 4$ (**c**), $n = 3$ (**d**) biologically independent experiments.

accumulated in CD34⁺ bone marrow cells from patients with DBA³⁴. Thus, the KU-mediated assembly of a catalytically inactive or T2609 phosphorylation-defective DNA-PK leads to rRNA processing defects. To identify the spectrum of cellular RNAs that are bound by KU and DNA-PKcs, we performed infrared crosslinking and immunoprecipitation³⁵ (irCLIP) of KU86 and DNA-PKcs (Extended Data Fig. 10a, b). Annotation of the reverse transcriptase (RT) stops across the transcriptome revealed the interaction profile of KU86 and DNA-PKcs (Extended Data Fig. 10c, e). Although most KU86 and DNA-PKcs RT stops mapped to introns of mRNA or non-coding RNAs (Extended Data Fig. 10c–f), they showed little overlap in this category (Extended Data Fig. 10e, f). The RT stops shared by KU86 and DNA-PKcs included known DNA-PK-interacting RNAs, such as *Terc* and *Neat1*³⁶ (Extended Data Fig. 10g, h). We also observed binding to non-coding RNAs with essential roles in ribosome biogenesis, including the 5'-ETS and U3 (Fig. 4a, b). Binding of DNA-PK to the 5'-ETS was sensitive to ActD (Extended Data Fig. 10i, j), which might contribute to the Pol I-dependent localization of DNA-PKcs and KU in the nucleolus (Extended Data Fig. 8a–c).

The U3 snoRNA can activate DNA-PK

DNA-PKcs and KU86 RT stops were biased towards the 5' end of U3, near the conserved hinge region (Fig. 4b) that is essential for 18S rRNA maturation^{4,32}. By contrast, DNA-PKcs and KU86 showed little preference towards the snoRNP domain of U3 (Fig. 4b). Correspondingly, DNA-PKcs and KU86 bound few and dissimilar snoRNAs, when compared to the snoRNA-binding protein DDX21³⁷ (Extended Data Fig. 11a, b). To understand whether KU86 drives the assembly of DNA-PK at U3, we mapped the DNA-PKcs and KU86 peaks to an in silico secondary structure (Fig. 4c) of U3³⁸. The DNA-PKcs and KU86 crosslinking sites are adjacent to one another in the same predicted stem-loop of U3

(U3-SL1; Fig. 4c), and within the DNase I footprint of DNA-PK³⁹. KU86 crosslinked mainly to the terminal hairpins (Fig. 4c). KU can assemble DNA-PK at DNA fragments with hairpin ends. Incubation of U3-SL1 with the native DNA-PK complex, purified from human cells, showed that KU bound to a substantial fraction of U3-SL1 (Extended Data Fig. 11c). Anti-KU86 antibodies supershifted the KU–U3-SL1 complex (Extended Data Fig. 11d). A mutated U3-SL1 hairpin failed to compete with wild-type U3-SL1 for KU86 binding (Extended Data Fig. 11e), suggesting that RNA structure is important for binding of KU. At higher concentrations of DNA-PK, KU was able to drive assembly of DNA-PK on U3-SL1 in the absence of ATP (Extended Data Fig. 11c). Moreover, U3-SL1 drove auto-phosphorylation of DNA-PK at the T2609 cluster (Extended Data Fig. 11f), but no efficient phosphorylation was detected at the T2056 cluster (Extended Data Fig. 11g). This RNA-based activity depended on DNA-PKcs kinase, as NU7441 or non-hydrolysable ATP blocked T2609 phosphorylation (Extended Data Fig. 11h–j). Similarly, U3-SL1 promoted baculovirus-purified human DNA-PK to undergo auto-phosphorylation at T2609, but not S2056 (Fig. 4d, Extended Data Fig. 11k). U3-SL1 activated DNA-PK to phosphorylate serine 15 on TP53 peptides, with efficiency comparable to the basal activity of isolated ATM (Fig. 4e). Notably, DNA-PK assembly and auto-phosphorylation at U3-SL1 is less efficient than with dsDNA (Extended Data Fig. 11f). We speculated that additional factors might be available to facilitate efficient DNA-PK assembly and activation by RNA.

RNA-dependent function of DNA-PK

Despite the accumulation of KU and DNA-PKcs in the nucleolus, young *Ku*^{-/-} or *DNA-PKcs*^{-/-} mice do not have deleterious translation defects. Using *DNA-PKcs*^{KD/KD} and *DNA-PKcs*^{3A/5A} mouse models, we have identified an unexpected role for DNA-PK during ribosome biogenesis and

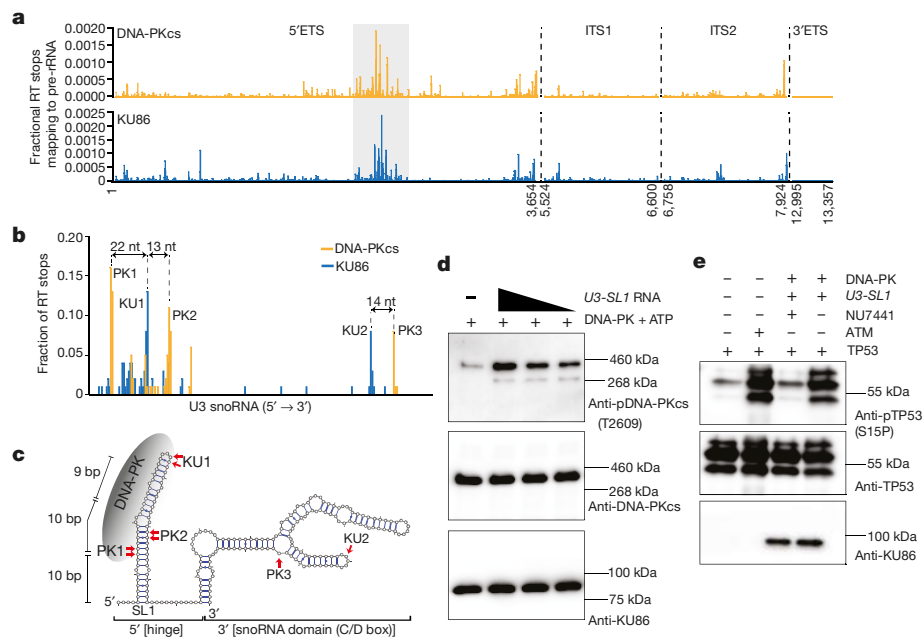


Fig. 4 | U3 snoRNA drives assembly, activation, and auto-phosphorylation of DNA-PK at the T2609 cluster. **a**, DNA-PKcs and KU86 RT stops that map to the four pre-rRNA introns; 5'ETS, ITS1, ITS2, and 3'ETS. irCLIP was performed in biological duplicate. **b**, Transcript normalized histogram of DNA-PKcs (orange) and KU86 (blue) irCLIP RT stops from DMSO-treated HeLa cells mapping to U3. **c**, In silico (mFold)-predicted secondary structure of U3. The top three and two peaks of DNA-PKcs and KU86, respectively, are annotated on the secondary structure. The length of the 5' end stem-loop is shown. **d**, Baculovirus-purified human DNA-PK in vitro kinase phosphorylation assay in the presence of

increasing amounts of U3-SL1. Western blots were performed with antibodies against DNA-PKcs phosphorylated at the T2609 cluster (top), total DNA-PK (middle), and KU86 (bottom). **e**, Baculovirus-purified human DNA-PK in vitro kinase phosphorylation assay against purified human TP53. Activated DNA-PK phosphorylates TP53 on serine 15. An in vitro ATM kinase assay was used as a control. Western blots were performed with antibodies against TP53 phosphorylated on serine 15 (top), total TP53 (middle) and KU86 (bottom). **d, e**, $n = 3$ biologically independent experiments.

demonstrated how an endogenous RNA in mammalian cells drives the assembly and activation of DNA-PK. The different effects of *DNA-PKcs*^{-/-} and *DNA-PKcs*^{KD/KD} on ribosome biogenesis mirror their effects on cNHEJ, where end-ligation is abrogated in *DNA-PKcs*^{KD/KD} cells but not in *DNA-PKcs*^{-/-} cells; this suggests that, once recruited to DNA or RNA, the kinase activity of DNA-PKcs might allosterically regulate the accessibility of DNA-PK-bound DNA or RNA to regulate repair or processing. Moreover, U3 triggers DNA-PKcs phosphorylation at T2609, which is critical for rRNA processing, but relatively dispensable for cNHEJ, implying that RNA- versus DNA-bound DNA-PK might undergo different conformational changes and are subject to different regulation. In this context, we propose that the inability of DNA-PK to regulate itself when assembled at U3 and/or other structured RNAs blocks SSU assembly and pre-rRNA processing, leading to translation defects that preferentially affect cell types and tissues with high demand for protein synthesis (for example, haematopoietic cells and ES cells). Although we focused on U3, our irCLIP analysis uncovered a myriad of other RNAs that could assemble DNA-PK, which will significantly expand the functional domain of DNA-PK. Finally, we noticed that in a *Tp53*-deficient background, *DNA-PKcs*^{KD/KD} and *DNA-PKcs*^{SA/SA} mice have different overall survival, which might be due to the differing effects of these mutations on specific RNAs or the added cNHEJ defects in *DNA-PKcs*^{KD/KD} mice. Further investigation will establish the relevance of DNA-PK assembly at RNA and how it differs from DNA, and whether they could both collaborate to promote DNA repair and RNA processing at highly transcribed regions⁴⁰.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2041-2>.

- Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
- Jiang, W. et al. Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining. *Mol. Cell* **58**, 172–185 (2015).
- Zhu, C. et al. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell* **109**, 811–821 (2002).
- Dragon, F. et al. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002).
- Alt, F. W., Zhang, Y., Meng, F. L., Guo, C. & Schwer, B. Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* **152**, 417–429 (2013).
- Gao, Y. et al. A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for KU in V(D)J recombination. *Immunity* **9**, 367–376 (1998).
- Kirchgesner, C. U. et al. DNA-dependent kinase (p350) as a candidate gene for the murine SCID defect. *Science* **267**, 1178–1183 (1995).
- Taccioli, G. E. et al. Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity* **9**, 355–366 (1998).
- Frank, K. M. et al. DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol. Cell* **5**, 993–1002 (2000).
- Gao, Y. et al. Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* **404**, 897–900 (2000).
- Nacht, M. et al. Mutations in the p53 and SCID genes cooperate in tumorigenesis. *Genes Dev.* **10**, 2055–2066 (1996).
- Crowe, J. L. et al. Kinase-dependent structural role of DNA-PKcs during immunoglobulin class switch recombination. *Proc. Natl Acad. Sci. USA* **115**, 8615–8620 (2018).
- Jacks, T. et al. Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
- Meek, K., Dang, V. & Lees-Miller, S. P. DNA-PK: the means to justify the ends? *Adv. Immunol.* **99**, 33–58 (2008).
- Davis, A. J., Chen, B. P. & Chen, D. J. DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA Repair (Amst.)* **17**, 21–29 (2014).
- Jiang, W. et al. Phosphorylation at S2053 in murine (S2056 in human) DNA-PKcs is dispensable for lymphocyte development and class switch recombination. *J. Immunol.* **203**, 178–187 (2019).
- Zhang, S. et al. Congenital bone marrow failure in DNA-PKcs mutant mice associated with deficiencies in DNA repair. *J. Cell Biol.* **193**, 295–305 (2011).
- Lee, B. S. et al. Functional intersection of ATM and DNA-PKcs in coding end joining during V(D)J recombination. *Mol. Cell Biol.* **33**, 3568–3579 (2013).

19. Houghtaling, S. et al. Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. *Genes Dev.* **17**, 2021–2035 (2003).
20. Williams, E. S. et al. Telomere dysfunction and DNA-PKcs deficiency: characterization and consequence. *Cancer Res.* **69**, 2100–2107 (2009).
21. Zhang, S. et al. Spontaneous tumor development in bone marrow-rescued DNA-PKcs(3A/3A) mice due to dysfunction of telomere leading strand deprotection. *Oncogene* **35**, 3909–3918 (2016).
22. Yoo, S. & Dynan, W. S. Characterization of the RNA binding properties of Ku protein. *Biochemistry* **37**, 1336–1343 (1998).
23. Chen, H. et al. Structural insights into yeast telomerase recruitment to telomeres. *Cell* **172**, 331–343.e313 (2018).
24. Pfingsten, J. S. et al. Mutually exclusive binding of telomerase RNA and DNA by Ku alters telomerase recruitment model. *Cell* **148**, 922–932 (2012).
25. Britton, S., Coates, J. & Jackson, S. P. A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *J. Cell Biol.* **202**, 579–595 (2013).
26. Narla, A. & Ebert, B. L. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115**, 3196–3205 (2010).
27. Signer, R. A., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49–54 (2014).
28. Khajuria, R. K. et al. Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. *Cell* **173**, 90–103.e119 (2018).
29. Unnikrishnan, I., Radfar, A., Jenab-Wolcott, J. & Rosenberg, N. p53 mediates apoptotic crisis in primary Abelson virus-transformed pre-B cells. *Mol. Cell. Biol.* **19**, 4825–4831 (1999).
30. Strauss, M. E. & Brandt, J. Attempt at preclinical identification of Huntington's disease using the WAIS. *J. Clin. Exp. Neuropsychol.* **8**, 210–218 (1986).
31. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
32. Barandun, J., Hunziker, M. & Klinge, S. Assembly and structure of the SSU processome—a nucleolar precursor of the small ribosomal subunit. *Curr. Opin. Struct. Biol.* **49**, 85–93 (2018).
33. Zhou, F. et al. Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. *J. Proteome Res.* **9**, 6242–6255 (2010).
34. Flygare, J. et al. Human RPS19, the gene mutated in Diamond-Blackfan anemia, encodes a ribosomal protein required for the maturation of 40S ribosomal subunits. *Blood* **109**, 980–986 (2007).
35. Zarnegar, B. J. et al. irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* **13**, 489–492 (2016).
36. Morchikh, M. et al. HEXIM1 and NEAT1 long non-coding RNA form a multi-subunit complex that regulates DNA-mediated innate immune response. *Mol. Cell* **67**, 387–399.e385 (2017).
37. Calo, E. et al. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253 (2015).
38. Parker, K. A. & Steitz, J. A. Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. *Mol. Cell. Biol.* **7**, 2899–2913 (1987).
39. Weterings, E., Verkaik, N. S., Brüggewirth, H. T., Hoeijmakers, J. H. & van Gent, D. C. The role of DNA dependent protein kinase in synopsis of DNA ends. *Nucleic Acids Res.* **31**, 7238–7246 (2003).
40. van Sluis, M. & McStay, B. A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev.* **29**, 1151–1163 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Mouse models

DNA-PKcs^{-/-}, *Xrcc4*^{+/-}, *DNA-PKcs*^{+/*KD*}, *DNA-PKcs*^{+/*PQR*}, *Ku70*^{+/-}, and *Tp53*^{+/-} alleles and mice have been described^{2,6,16,41–43}. The DNA-PKcs PQR allele substitutes five serine residues within the S2056 cluster (S2053 in mouse) with alanine (Extended Data Fig. 2a). The DNA-PKcs-5A mutation converts all five threonines in the T2609 cluster of exon 58 of DNA-PKcs to alanine (Extended Data Fig. 2b–d). The targeting arms (5' and 3' arms) were generated via PCR, cloned into pGEMT and sequence verified, before being subcloned into the pEMC-neo targeting vector. The ~500-bp fragment containing all the mutations was synthesized (Genewiz) and integrated into the 3' arm (4.4 kb) to generate the targeting vector (Extended Data Fig. 2b, c) with inverted neo-resistance (NeoR) cassettes flanked by a pair of FRT sites. An SmaI-linearized targeting plasmid was electroporated into mouse ES cells (129/Sv background) and NeoR-resistant clones were isolated and screened via Southern blotting with SpeI digestion and a 5' probe (amplified via primers 5'-TCATTAAGGCCTCGCAATC-3' and 5'-TGTTGGAGAAGGATGGGAAA-3'; germline, 14.7 kb; targeted, 5.8 kb). The successfully targeted clones were validated via Southern blotting with a NeoR probe for single integration and sequenced to verify the desired mutation, before being injected for germline transmission. The resultant DNA-PKcs^{+/*SAN*} chimeras (N for neo positive) were crossed with constitutively FLIPase-expressing *Rosa26a*^{FLIP/FLIP} mice (Jax stock number: 003946, also in 129/Sv background) to remove the NeoR cassette and allow the expression of DNA-PKcs-5A. Genotyping was performed with primers (5'-GCAGATCTCTGTGAATTTATGACC-3' and 5'-CCAAGAAAACTAACGAAAAATGC-3'). The product corresponding to the wild-type allele is 172 bp and the product corresponding to the 5A allele is 244 bp (including an FRT site). Complete blood counts were measured by the Institute of Comparative Medicine at Columbia University using Genesis (Oxford Science Inc.) and histological analyses were performed by the histopathological core of Herbert Irving Comprehensive Cancer Center (HICCC) at Columbia University. All animal work was conducted in an on-site, pathogen-free facility in Columbia University and following procedures pre-approved by the Institutional Animal Care and Use Committee (IACUC) at Columbia University Medical Center. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Flow cytometry analyses of haematopoietic cells and lymphocytes

Flow cytometry analyses of haematopoietic cells and lymphocytes were performed using previously published methods^{2,44,45} with minimal colour modifications. In brief, single-cell suspensions of bone marrow or fetal liver cells were filtered through a 70-µm nylon cell strainer (Falcon, 352350). Splenocytes were treated with red blood cell lysis buffer before staining. For HSPC analysis, bone marrow or fetal liver cells were first stained with a cocktail of biotin labelled primary antibodies containing mouse specific lineage markers: anti-CD4 (Biolegend, 100404), anti-CD8a (Biolegend, 100704), anti-CD3e (Biolegend, 100304), anti-CD5 (BD Bioscience, 553018), anti-B220 (Biolegend, 103204), anti-Ly6G/Ly6C (Gr-1) (Biolegend, 108404) and anti-TER119 (Biolegend, 116204) at 4 °C for 1 h, before being washed and stained with a cocktail of directly conjugated antibodies including FITC anti-mouse CD34 (eBioscience, 11-0341-82), PE anti-mouse Ly-6A/E (Sca-1) (Biolegend, 108108), PE/Cy7 anti-mouse CD16/32 (Biolegend, 101318), APC anti-mouse CD117 (c-Kit) (135108), APC/Cy7 Streptavidin (Biolegend, 405208) and Brilliant Violet 510 anti-mouse CD41 (Biolegend, 133923) at 4 °C for 1.5 h. For lymphocyte and peripheral haematopoietic cells analyses, cells from the bone marrow, spleen, or thymus were stained with B cell cocktail (FITC anti-mouse CD43, Biolegend, 553270; PE goat anti-mouse IgM, Southern Biotech, 1020-09; PE-cyanine5 anti-Hu/CD45R (B220),

eBioScience 15-0452-83; and APC anti-mouse TER119, Biolegend 116212); T cell cocktail (PE rat anti-mouse CD4, Biolegend 557308; FITC anti-mouse CD8a, Biolegend 100706; PE/Cy5 anti-mouse CD3e, eBioscience 15-0031-83; and APC anti-mouse TCRβ, BD Pharmingen 553174); myeloid cocktail (FITC anti-mouse CD11b, BD Pharmingen 553310; PE rat anti-mouse CD19, BD Pharmingen 557399; PE/Cy5 anti-mouse CD3e, eBioscience 15-0031-83; and APC anti-mouse Ly6G/Ly6C (Gr-1), Biolegend 108412) and/or erythroid cocktail (PE-cyanine5 anti-Hu/CD45R (B220) (as above); PE/Cy7 anti-mouse TER119, Biolegend, 116222; APC anti-mouse Ly6G/Ly6C (Gr-1) (as above); Pacific Blue anti-mouse Thy1.2, Biolegend 140306; and Brilliant Violet 510 anti-mouse CD71, Biolegend 113823), according to specific cell type containing. To analyse the double-negative T cells in the thymus, total thymocytes were stained for FITC-anti CD25 (BD Pharmingen, 1595-02S), and APC anti-mouse CD44 (eBioscience, 17-0441-82) together with a collection of PE-conjugated antibodies to remove all other cell lineages (including anti-mouse CD4, CD8a (BD Pharmingen, 553032), CD19, TCRγδ (eBioscience, 12-5711-82)) and PE/Cy7 anti-mouse TER119. The flow cytometry data were collected on either LSR II (BD) with BD FACSDiva software, or on a FACSCalibur (BD) with BD CellQuest Pro. All flow cytometry data were analyzed using FlowJo V10.

Cell culture and cell line derivation

The *DNA-PKcs*^{+/-} and *DNA-PKcs*^{-5A/*SA*} MEFs were derived from E13.5–E14.5 embryos obtained from crosses between *DNA-PKcs*^{+/*SA*} parents using standard procedures^{2,46}. The *DNA-PKcs*^{+/-} and *DNA-PKcs*^{-5A/*SA*} ES cells were derived by super-ovulating 3–4-week-old *DNA-PKcs*^{+/*SA*} female mice mated with adult male *DNA-PKcs*^{+/*SA*} mice. The inner cell masses from E3.5 blastocysts were then isolated and expanded for genotyping. The newly derived ES cells and controls were cultured on irradiated (30 Gy) fibroblast feeders using standard ES cell medium (DMEM, Gibco 12430-062; 15% fetal bovine serum, Hyclone, SH30071.03; 1 × MEM non-essential amino acids, Gibco 11140-050; 1 mM sodium pyruvate, Gibco 11360-070, 2 mM L-glutamine, Gibco 25030-081; 120 µM 2-mercaptoethanol, Fisher 034461-100; penicillin/streptomycin, Gibco 15070-63 and leukaemia inhibition factor (LIF, provided by C.-S. (Victor) Lin)). The v-ABL kinase-transformed B cell lines were derived from EµBCL2⁺ transgenic mice with the indicated genotype as previously described⁴⁷. In brief, single-cell suspension from total bone marrow (from 2–5-week-old mice) or fetal liver was infected with retrovirus encoding p120 minimal v-ABL kinase⁴⁷. Clonal outgrowth was isolated and expanded for the next 6–8 weeks. CRISPR-Cas9-mediated deletion of *Ku80* was performed as previously described².

The *Fancd2*^{-/-} MEFs⁷ were generously provided by A. Smogorzewska (Rockefeller University). The v-ABL kinase-transformed *DNA-PKcs*^{-3A/*3A*} B cell lines^{17,18} were generously provided by B. Sleckman (Cornell University).

Measurement of protein synthesis using OPP

For in vitro analyses of cell lines, 5 × 10⁶ v-ABL kinase-transformed B cells were plated in each well of a 24-well plate. ATM or DNA-PK kinase inhibitors were added 17 h before the cells were treated with 5 µM OPP (ThermoFisher) for 30 min. To prevent protein syntheses (establishing a negative control), 100 µg/ml cycloheximide (CHX; Sigma) was added at the same time as OPP in a subset of samples. For all the translation experiments, the azide-alkyne cycloaddition was performed using the OPP Protein Synthesis Assay Kit (ThermoFisher Scientific). In brief, the cells were washed with ice-cold PBS, fixed with 3.7% formaldehyde in PBS for 15 min at 25 °C, permeabilized with 0.5% Triton X-100/PBS for 15 min at 25 °C, and then subjected to azide-alkyne cycloaddition and detection with azide-conjugated Alexa Fluor 488. The data were collected using FACSCaliber (BD Biosciences for cell lines) or BD FACSDiva (for haematopoietic cells) and analysed using FlowJo V10.

Translation analyses of haematopoietic cells from mice were conducted as previously described with minor modifications²⁷. In brief,

Article

fresh isolated marrow and splenocytes were cultured for 1 h in RPMI 1640 complete medium (Gibco 11875) with 15% FBS containing 5 μ M OPP before staining of live haematopoietic stem cells (see above with minor modification of the colour to accommodate FITC-OPP). Specifically, Alexa Fluor 700 rat anti-mouse CD34 (eBioscience, 560518) and Ly-6A/E (Sca-1) and Pacific Blue anti-mouse Ly-6A/E (Sca-1) (Biolegend, 108120) were used for HSPC staining. The relative levels of protein synthesis between different haematopoietic populations obtained from transient in vitro incubation of bone marrow from 2-week-old *DNA-PKcs*^{+/+} mice (Extended Data Fig. 7d) agreed with prior data obtained with in vitro and in vivo injection of OPP in adult mice²⁷. Specifically, among the HSPC and lineage specific progenitors, the highest protein translation was found in megakaryocytes and erythroid progenitors (MEPs), followed by common myeloid progenitors (CMPs). HSPCs (Lin[−]SCA1⁺c-KIT⁺; LSK) have relatively low protein synthesis²⁷ (Extended Data Fig. 7d). To offset the prohibitory cost of in vivo OPP injection, we measured global translation with transient (1 h) in vitro incubation.

Northern blot analysis of rRNA processing

Total RNA was extracted using TRIzol (Life Technology) and about 5 μ g total RNA was analysed for each sample. Oligo probes measuring 18S and 5.8S rRNA processing situated at the 3' boundary of the mature 18S or 5.8S species were designed on the basis of previous publications^{48,49} (Supplementary Table 1).

Immunofluorescence

U2OS cells or mouse ES cells were seeded on gelatin-treated coverslips 48 h before staining and stained as previously described²⁵. In brief, the cells were washed with 1 \times PBS, when indicated, incubated with CSK buffer²⁵ (10 mM Pipes with pH 7.0, 100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂ and 0.7% Triton X-100) for 3 min at room temperature for pre-extraction, then washed and fixed with 4% paraformaldehyde (PFA) in PBS for 25 min at 25 °C. Before staining, the cells were permeabilized with 0.1% Triton X-100/PBS for 10 min and blocked with 3% BSA for 1 h at 25 °C. Fixed cells were then incubated with primary antibodies in 3% BSA for 1 h at 25 °C, including mouse anti-human KU86 (ThermoFisher, MA5-12933, 1:100), rabbit anti-human DDX21 (Novus, NB100-1718, 1:500) or anti-DNA-PKcs (ThermoFisher, Ab-4 (cocktail)), followed by fluorophore-conjugated secondary antibodies (Alexa Fluor 488-conjugated anti-rabbit, Alexa Fluor 594-conjugated anti-rabbit, and cyanine3-conjugated anti-mouse, Invitrogen, 1:500) for 1 h at room temperature. All images were captured on a Nikon A1RMP confocal microscope with a 63 \times objective.

Cell cultures

HeLa and U2OS cells were cultured in DMEM (plus 10% FBS) and maintained under standard tissue-culture conditions unless otherwise specified. For actinomycin D (Sigma) treatments, 1 μ g/ml was added to the culture medium for 1 h before harvesting for downstream assays. For immunofluorescence studies, cells were fixed with 4% paraformaldehyde, unless otherwise specified, and stained with the indicated antibodies. All cell lines used in this study were mycoplasma-free.

Infrared crosslinking and immunoprecipitation

irCLIP was performed as described³⁵. HeLa cells were grown as described above and UV crosslinked with a total of 0.35 J/cm². Whole-cell lysates were generated in CLIP lysis buffer (50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40, 0.2% Triton X-100, 0.5% *N*-lauroylsarcosine) and briefly sonicated using a probe-tip Branson sonicator to solubilize chromatin. Each experiment was normalized to total protein amount, typically 1 mg, and partially digested with RNase A (ThermoFisher Scientific, EN0531) for 10 min at 37 °C and quenched on ice. DNA-PKcs (Bethyl, A303-967A), KU80 (Bethyl, A302-627A) and IgG (ThermoFisher Scientific, 02-6102) IPs were performed with 15 μ g antibody and 50 μ l Protein G Dynabeads (ThermoFisher Scientific) for 8

h at 4 °C on rotation. Samples were washed for 1 min at 25 °C each time in 1 ml of the following buffers sequentially: 1 \times high stringency buffer (15 mM Tris-HCl, pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 120 mM NaCl, 25 mM KCl), 1 \times high salt buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 1 M NaCl), 2 \times NT2 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM MgCl₂, 0.05% NP-40). After the NT2 wash, RNA-protein complexes were dephosphorylated with T4 PNK (NEB) for 45 min in an Eppendorf Thermomixer at 37 °C with repeated cycle of 15 s at 1,400 rpm followed by 90 s of rest in a 30- μ l reaction system of pH 6.5 containing 10 units T4 PNK, 0.1 μ l SUPERase-IN (ThermoFisher Scientific), and 6 μ l PEG-400 (16.7% final). Dephosphorylated RNA-protein complexes were then rinsed once with NT2 buffer and 3'-end ligated with T4 RNA Ligase 1 (NEB) overnight in an Eppendorf Thermomixer at 16 °C with a repeated cycle of 15 s at 1,400 rpm followed by 90 s of rest in a 60- μ l reaction containing 10 units T4 RNA Ligase, 1.5 pmol pre-adenylated-IR800-3'biotin DNA-adaptor, 0.1 μ l SUPERase-IN, and 6 μ l PEG400 (16.7% final). The following day, samples were rinsed once again with 500 μ l NT2 buffer and resuspended in 30 μ l NT2 buffer containing 20 mM DTT and 1 \times LDS (ThermoFisher Scientific). Samples were then heated to 75 °C for 10 min, and released RNA-protein complexes were separated on 4–12% Bis-Tris SDS-PAGE gel (1.0 mm \times 12 well) at 200 V for 45 min. Resolved RNP complexes were wet-transferred to nitrocellulose at 550 mA for 45 min at 4 °C.

Nitrocellulose membranes were imaged using an Odyssey CLx scanner (LiCor), RBP-RNA complexes were excised using scalpels, and RNA was recovered by adding 0.1 ml proteinase K reaction buffer (100 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.2% SDS and 5 μ l of 20 mg/ml proteinase K (ThermoFisher Scientific)). Proteins were digested for 60 min at 50 °C in an Eppendorf Thermomixer. Next, 200 μ l saturated phenol-chloroform pH 6.7 was added to each tube and incubated for 10 min at 37 °C in an Eppendorf Thermomixer at 1,400 rpm. Tubes were briefly centrifuged, and the entire contents were transferred to a 2-ml Heavy Phase Lock Gel tube (SPrime, 2302830). Samples were centrifuged for 2 min at >13,000 rpm. The aqueous layer was re-extracted with 1 ml chloroform (inverting 10 times to mix; no vortexing) in the same 2-ml Phase Lock Gel tube and centrifuged for 2 min at >13,000 rpm. The aqueous layer was then transferred to a new 2-ml Heavy Phase Lock Gel tube and extracted again with an additional 1 ml chloroform. After 2 min centrifugation at >13,000 rpm, the aqueous layer was transferred to a siliconized 1.5-ml tube and precipitated overnight at −20 °C by addition of 10 μ l 5 M NaCl, 3 μ l linear polyacrylamide (ThermoFisher Scientific) and 0.8 ml 100% ethanol. RNA fragments were pelleted at >13,000 rpm for 45 min at 4 °C, washed once with 1 ml ice-cold 75% ethanol and air dried.

RNA pellets were resuspended in 12 μ l water by adding 1 μ l of 3 μ M cDNA and 1 μ l of 10 mM dNTPs, and heated to 70 °C for 5 min before being rapidly cooled to 4 °C. cDNA Master Mix (4 μ l 5 \times Super Script IV (SSIV) buffer, 1 μ l 100 mM DTT, 1 μ l SSIV, 6 μ l total) was added to the annealed RNA and incubated for 30 min at 55 °C. cDNA-RNA hybrids were captured by addition of 5 μ l of MyOne Streptavidin C1 Dynabeads (ThermoFisher Scientific) that had been rinsed and suspended in 50 μ l biotin-IP buffer (100 mM Tris, pH 7.5, 1 M NaCl, 1 mM EDTA, 0.1% Tween), and subjected to end-over-end rotation for 45 min at room temperature. Beads were placed on a 96-well magnet and washed twice with 100 μ l biotin-IP buffer and 100 μ l ice-cold 1 \times PBS sequentially. Beads were resuspended in 10 μ l cDNA elution buffer (8.25 μ l water adding 1 μ l of 1 μ M P3 short oligo and 0.75 μ l of 50 mM MnCl₂) and heated to 95 °C for 10 min, ramped at 0.1° s^{−1} to 60 °C until the end of the experiment. cDNA was circularized for 2 h at 60 °C in 5 μ l circularization reaction buffer (3.3 μ l water, 1.5 μ l 10 \times Circligase-II buffer, and 0.5 μ l Circligase-II (Epicentre)), and then purified with 30 μ l AMPure XP beads (Beckman Coulter) and 75 μ l isopropanol. Samples were incubated for 20 min at 25 °C, washed twice with 100 μ l 80% ethanol, air dried for 5 min, and eluted in 14 μ l water. Elution took place at 95 °C for 3 min and samples

were then immediately transferred to a 96-well magnet. Eluted cDNA was transferred to a new PCR tube containing 15 µl 2× Phusion HF-PCR Master Mix (NEB), 0.5 µl of 30 µM P3/P6 PCR1 oligo mix and 0.5 µl of 15× SYBR Green I (ThermoFisher Scientific). Real-time quantitative PCR was performed as: 98 °C for 2 min, 15 cycles of 98 °C for 15 s followed by 65 °C for 30 s, and 72 °C for 30 s, and eventually with data acquisition set to the 72 °C extension. PCR1 reactions were purified by adding 4.5 µl isopropanol and 54 µl AMPure XP beads before incubation for 10 min. Beads were washed once with 80% ethanol, dried for 5 min, and eluted in 15 µl water. Illumina flow cell adaptors were added by adding 15 µl 2× Phusion HF-PCR Master Mix and 0.4 µl P3solexa/P6solexa oligo mix and amplified as: 98 °C for 2 min, 3 cycles of 98 °C for 15 s followed by 65 °C for 30 s, and 72 °C for 30 s seconds. Final libraries were purified by addition of 48 µl AMPure XP beads and incubation for 5 min. Beads were washed twice with 70% ethanol, dried for 5 min, and eluted in 20 µl water. Libraries (1–2 µl) were quantified using an HS-DNA Bioanalyzer. Samples were deep sequenced on the Illumina NextSeq machine with 75-bp cycle run of single-end and no index high-output sequencing.

Analysis of irCLIP data

The irCLIP data were processed using the FAST-iCLIP pipeline (<https://github.com/ChangLab/FAST-iCLIP/tree/lite>). PCR duplicates were removed using unique molecular identifiers (UMIs) in the RT primer region. Adaptor and barcode sequences were trimmed, and reads were mapped step-wise to repetitive and then non-repetitive (GRCh38) genomes. Specific parameters used for the FAST-iCLIP pipeline are as follows: -f18 (trims 17 nt from the 5' end of the read), -l16 (includes all reads longer than 16 nt), -bm29 (minimum MAPQ score from bowtie2 of 29 is required for mapping; unique mapping only), and -tr2,3 (repetitive genome) and -tn2,3 (non-repetitive genome) RT stop intersection (n,m ; where n = replicate number and m = number of unique RT stops required per n replicates). Using the -tr/tn2,3 parameters, a minimum of six RT stops is required to support any single nucleotide identified as crosslinking site. For gene ontology analysis, the DAVID⁵⁰ tool was employed (<https://david.ncicrf.gov/tools.jsp>). Peaks of RT stops were called on the biological replicated intersection of RT stop positions using iCount peaks (<http://icount.readthedocs.io/en/latest/>). The command line was as follows: iCount peaks gencode.v21.annotation.segment.gtf RTstop_input.bed Out_iCpeaks.bed -scores Out_iCpeaksScores.tsv. Regions from 'Out_iCpeaks.bed' were then annotated using HOMER (<http://homer.ucsd.edu/homer/>) using the following command: annotatePeaks.pl Out_iCpeaks.bed hg38> Out_iCpeaks_hg38_HOMERanno.txt -annStats Out_iCpeaks_hg38_HOMERanno_stats.txt.

Comprehensive identification of RNA binding proteins by mass spectrometry

U3 snoRNA targeting probes were designed using online tools available at <https://www.biosearchtech.com/stellaris>, with a repeat masking setting of 3 and even coverage of the whole transcript. Full probe sequences are available in Supplementary Table 2. Oligos were synthesized with 3' biotin-TEG modification at Stanford Protein and Nucleic Acid Facility (panoligo@stanford.edu).

ChIRP-MS was performed largely as described³¹. Five 15-cm dishes of cells (~200 million cells) were used per ChIRP-MS experiment. Cell culture and ActD treatments were performed as described above. Medium was aspirated and cells were rinsed once with 5 ml of 1× PBS. PBS was then aspirated, and 10 ml of 3% formaldehyde in 1× PBS was added to each dish. Cells were scraped off the 15-cm dishes during chemical crosslinking, which took place for a total of 30 min at 25 °C and was stopped by the addition of glycine to a final concentration of 125 mM for 5 min at 25 °C. Crosslinked cells were pelleted at 2,000 rpm for 5 min, supernatant was discarded, and cells frozen at -80 °C for storage. Lysate was generated by resuspending cell pellets in 1 ml lysis buffer (50 mM Tris-HCl pH 7.0, 10 mM EDTA, 1% SDS) per 100 mg of cell pellet weight (~100 µl pellet volume). Lysates were sonicated using a focused

ultrasonicator (Covaris, E220) until the average RNA length was 500 nucleotides as determined by agarose gel analysis and stored at -80 °C. Stored lysates were thawed on ice and prepared for pre-clearing and RNase A treatment (control sample). The experimental samples were precleared by adding 30 µl washed MyOne C1 beads per ml of lysate at 37 °C for 30 min on rotation. To control sample preclearing, RNA digestion was achieved by adding 30 µl washed MyOne C1 beads per ml lysate and 1/1,000 the volume of 10 mg/ml RNase A (Sigma) and incubating at 37 °C for 30 min on rotation. Preclearing beads were collected twice from lysate using a magnetic stand; for this and all subsequent magnetic stand steps we allowed >1 min of separation before removing any supernatant. Next, 2 ml of ChIRP hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris-HCl pH 7.0, 1 mM EDTA, 15% formamide; made fresh) and 2.5 µl of 100 µM ChIRP Probe Pools were added per ml of lysate. ChIRP Probe Pools (Supplementary Table 1) were comprised of seven antisense oligos (final concentration of 16 µM of each probe). Hybridization took place on rotation for 16 h at 37 °C. Subsequently, 250 µl of washed MyOne C1 beads per ml of lysate were added to each sample and incubated on rotation for 45 min at 37 °C. Enriched material was collected on the beads with a magnetic stand, and beads were washed 5 × 2 min in 1 ml ChIRP Wash Buffer (2 × NaCl-sodium citrate (SSC, ThermoFisher Scientific), 0.5% SDS) at 37 °C. To elute enriched proteins, beads were collected on a magnetic stand, resuspended in ChIRP biotin elution buffer (12.5 mM biotin, 7.5 mM HEPES, pH 7.9, 75 mM NaCl, 1.5 mM EDTA, 0.15% SDS, 0.075% sarkosyl, and 0.02% Na-deoxycholate), and mixed at 25 °C for 20 min on rotation and at 65 °C for 15 min shaking. Eluent was transferred to a fresh tube, and beads were eluted again. The two eluents were pooled (~1,200 µl), and residual beads were removed again using the magnetic stand. Twenty-five per cent total volume (300 µl) of trichloroacetic acid was added to the clean eluent and vortexed, and then samples were placed at 4 °C overnight for precipitation. The next day, proteins were pelleted at 21,000 relative centrifugal force (rcf) at 4 °C for 45 min. Supernatant was carefully removed, and protein pellets were washed once with ice-cold acetone. Samples were spun at 21,000 rcf at 4 °C for 5 min. Acetone supernatant was removed and tubes were briefly centrifuged again and, after removal of residual acetone, were left to air-dry on a bench-top. Proteins were then solubilized in 1× LDS buffer in NT2 with 20 mM DTT and boiled at 95 °C for 30 min with occasional mixing for reverse-crosslinking.

Protein samples were size-separated on bis-tris SDS-PAGE gels (Bio-Rad), and the gel was fixed and stained with the Colloidal Blue Staining Kit (ThermoFisher Scientific) as per the manufacturer's instructions. Each ChIRP-MS experiment was cut into seven slices from the SDS-PAGE and prepared independently. Gel slices were prepared for mass spectrometry by rinsing sequentially in 200 µl HPLC-grade water, 100% acetonitrile (ACN, ThermoFisher Scientific) and 50 mM ammonium bicarbonate (AmBic). Samples were reduced by adding 200 µl of 5 mM DTT in 50 mM AmBic and incubating at 65 °C for 35 min. The reduction buffer was discarded, and samples were cooled to room temperature. Alkylation was achieved by adding 200 µl of 25 mM iodoacetamide in 50 mM AmBic for 20 min at 25 °C in the dark. The alkylation buffer was discarded, samples were rinsed once in 200 µl 50 mM AmBic, and then they were washed twice for 10 min each in 200 µl of freshly prepared 50% ACN in 50 mM AmBic. After each wash, the supernatant was discarded, and after all washes, samples were dried for 3 h using a SpeedVac. Once dry, the proteins were digested by adding 100 ng trypsin in 200 µl of 50 mM AmBic for 16 h at 37 °C. Samples were subsequently acidified by adding formic acid to a final concentration of 2.5% and incubating at 37 °C for 45 min. Finally, samples were desalted using HyperSep Filter Plates with a 5–7-µl bed volume (ThermoFisher Scientific) following the manufacturer's instructions. Samples were eluted three times in 100 µl 80% ACN in 2.5% formic acid, dried on a SpeedVac, and resuspended in 10 µl 0.1% formic acid for mass spectrometry analysis. Desalted peptides were analysed by online capillary

nanoliquid chromatography with tandem MS (nanoLC–MS/MS). Samples were separated using a 20-cm reversed phase column fabricated in-house (100 µm inner diameter, packed with ReproSil–Pur C18-AQ 3.0 µm resin (Dr. Maisch GmbH)) that was equipped with a laser-pulled nanoelectrospray emitter tip. Peptides were eluted at a flow rate of 400 nl/min using a two-step linear gradient of 2–25% buffer B in 70 min and 25–40% B in 20 min (buffer A: 0.2% formic acid and 5% DMSO in water; buffer B: 0.2% formic acid and 5% DMSO in acetonitrile) in an Eksigent ekspert nanoLC-425 system (AB Sciex). Peptides were ionized with electrospray ionization into an Orbitrap Elite Hybrid Ion Trap–Orbitrap Mass Spectrometer (ThermoFisher Scientific). Instrument method parameters were as follows: MS1 resolution, 60,000 at 400 *m/z*; scan range, 340–1,600 *m/z*. The top 20 most abundant ions were subjected to collision-induced dissociation with a normalized collision energy of 35%, activation *q* 0.25, and precursor isolation width 2 *m/z*. Dynamic exclusion was enabled with a repeat count of 1, a repeat duration of 30 s, and an exclusion duration of 20 s.

FASTA sequences of the human proteome (Uniprot: UP000005640) were downloaded and used to search the data files (.raw) using Byonic (Protein Metrics) with the following parameters: semi-specific cleavage specificity at the C-terminal site of R and K allowing for 2 missed cleavages. Mass tolerance was set at 12 ppm for MS1s, 0.4 for MS2s. Methionine oxidation, asparagine deamidation, and N-term acetylation were set as variable modifications. Cysteine carbaminomethylation was set as a fixed modification. Peptide hits were filtered using a 1% FDR. To visualize ChIRP–MS protein hits, paired Uniprot ID and ChIRP–MS enrichment values were imported into Cytoscape using the GeneMANIA extension⁵¹. Connections between hits visualized were selected to display only direct protein–protein interactions. Each protein hit was coloured by the enrichment value ($\log_2 \text{ChIRP} / \log_2 \text{RNase}$).

ChIRP–qRT–PCR

Cells were grown, crosslinked, and sonicated as described above. After sonication, 1% of the lysate was removed and saved as an ‘input’ sample. Lysates were again processed as above for preclearing, hybridization, MyOneC1 capture, and bead washing. After washing, 1% of each sample was removed as an ‘enriched’ fraction. Enriched fractions were collected while the MyOneC1 beads were fully resuspended in ChIRP wash buffer. The input and enriched samples were brought to 95 µl in ChIRP PK buffer (10 mM Tris–HCl pH 7.0, 100 mM NaCl, 1 mM EDTA, 0.2% SDS) and to this was added 5 µl of 20 mg/ml proteinase K. Protein was digested while shaking at 55 °C for 45 min. RNA was extracted by adding 500 µl TRIzol (ThermoFisher Scientific), incubating at 55 °C for 5 min, and then adding 100 µl chloroform. After mixing samples by vortexing for 7 s each, samples were incubated at 25 °C for 5 min and then spun at 12,000 rpm at 4 °C for 15 min. The aqueous layer was carefully removed from each sample, mixed with two volumes of 100% ethanol, and purified using an RNA Clean & Concentrator-25 (Zymo Research) per the manufacturer’s instructions. All RNA samples were DNase-treated with the Turbo DNA-Free kit (ThermoFisher Scientific). The cDNA was generated using SuperScript VILO (ThermoFisher Scientific) according to the manufacturer’s instructions. The qPCR analyses were performed on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). All primers used are shown in Supplementary Table 2.

Secondary structure folding of U3 snoRNA

The human U3 snoRNA sequence was obtained (NR_006880.1) and secondary structure folded using the mFold web server⁵² using default settings. A Vienna file (.b) was exported for the folded structure and visualized in VARNA⁵³.

Electrophoretic mobility shift assay

The stem-loop 1 region of U3 was synthesized with a T7 polymerase promoter and subsequently in vitro transcribed (IVT) using the MEGascript

T7 Transcription Kit (Thermo Fischer Scientific) per the manufacturer’s protocol to produce IVT–U3–SL1. To fluorescently label IVT–U3–SL1, periodate labelling of RNA 3’ ends was performed using Cy7-amine (Lumiprobe) as the dye. To perform the electrophoretic mobility shift assay (EMSA) reaction, IVT RNA was first denatured at 75 °C for 5 min, cooled rapidly on ice for 3 min, and then added to 1× EMSA buffer (10 mM HEPES pH 7.5, 20 mM KCl, 1 mM MgCl₂, and 1 mM DTT) for 5 min at 37 °C to allow the RNA to refold. For reactions that contained non-labelled competitor RNA or DNA, these were added at the refolding step. Non-labelled IVT–U3–SL1 and PCR DNA of U3–SL1 was added at 1:1 or 5:1 as unlabelled:labelled nucleic acids. For the DNA competitor, each end of the dsDNA was considered independently in the molar calculations. Finally, to each reaction we added 1 µl of 50% glycerol and recombinant DNA–PK holoenzyme (Promega), and water to 10 µl. EMSA reactions were incubated at 25 °C for 30 min and then run directly in a 4% native PAGE at 100 V. After electrophoresis, gels were directly scanned on a LiCor Odyssey in the 800 channel.

In vitro DNA–PK kinase reactions

IVT RNA, dsDNA, and DNA–PK holoenzyme were prepared as described above for EMSA analysis. Kinase reactions were assembled with 1 µl recombinant DNA–PK holoenzyme (Promega) and titrated amounts of nucleic acids (RNA, DNA, or none). To test for the dependence on ATP hydrolysis, 200 µM ATP, no ATP, or 200 µM adenosine-5’-[(β,γ)-methylene]triphosphate (AppCp) was added. When indicated, the DNA–PKcs inhibitor NU7441 was added to a final concentration of 1 µM. All reactions occurred for 60 min at 25 °C and were subsequently assayed by western blotting. The baculovirus-purified human DNA–PK (Thermo PV5866), TP53 (Millipore 23-034) and active human ATM (Millipore 14-933) were used in the assays. The kinase reactions were performed in 1× kinase buffer (Thermo PV3189) with 100 µM ATP and 1.5 µl recombinant DNA–PK holoenzyme or 1.0 µl ATM in the presence of TP53. All reactions occurred for 60 min at 25 °C and were subsequently assayed by western blotting.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

irCLIP data are available via the Gene Expression Omnibus (GEO) under the accession number GSE109026. All uncropped blots are provided in Supplementary Fig. 1. Data underlying the graphical representations used in the figures, including all experiments presenting data from animal models, are provided as SourceData. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.

Code availability

The FAST-iCLIP software is freely available at <https://github.com/ChangLab/FAST-iCLIP/tree/lite>.

- Gao, Y. et al. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891–902 (1998).
- Gu, Y., Jin, S., Gao, Y., Weaver, D. T. & Alt, F. W. Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. *Proc. Natl Acad. Sci. USA* **94**, 8076–8081 (1997).
- Symonds, H. et al. p53-dependent apoptosis suppresses tumor growth and progression in vivo. *Cell* **78**, 703–711 (1994).
- Avagyan, S. et al. Hematopoietic stem cell dysfunction underlies the progressive lymphocytopenia in XLF/Cernunnos deficiency. *Blood* **124**, 1622–1625 (2014).
- Liu, X., Shao, Z., Jiang, W., Lee, B. J. & Zha, S. PAXX promotes KU accumulation at DNA breaks and is essential for end-joining in XLF-deficient mice. *Nat. Commun.* **8**, 13816 (2017).
- Li, G. et al. Lymphocyte-specific compensation for XLF/Cernunnos end-joining functions in V(D)J recombination. *Mol. Cell* **31**, 631–640 (2008).
- Bredemeyer, A. L. et al. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466–470 (2006).

48. Wang, M., Parshin, A. V., Shcherbik, N. & Pestov, D. G. Reduced expression of the mouse ribosomal protein Rpl17 alters the diversity of mature ribosomes by enhancing production of shortened 5.8S rRNA. *RNA* **21**, 1240–1248 (2015).
49. Wang, M. & Pestov, D. G. 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. *Nucleic Acids Res.* **39**, 1811–1822 (2011).
50. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
51. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
52. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
53. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).

Acknowledgements We thank B. Sleckman and A. Smogorzewska for providing the control cell lines; W. Gu and T. Li for helping with Tp53 analyses in DNA-PKcs mouse models; R. Baer for commenting on the manuscript; L. Ding for discussion on hematopoietic stem cell analyses; T. Xiang and L. Berchowitz for assistance with the ribosomal and nucleoli analyses; C. S. Lin for germline injection and ES cell derivation; E. Zhang and S. Kaplan for help with telomere analyses; L. Zhang and the Elias laboratory for assistance running samples for MS analysis; J. Collier, D. Wagh, and the Stanford Functional Genomics Facility for assistance running samples for deep sequencing; and the Swanson Biotechnology Center at the Koch Institute for Integrative Cancer Research, especially the MIT BioMicro Center. Owing to space constraints, we often cited reviews rather than original publications. We apologize to the colleagues whose original works were not cited here. R.A.F. is supported by the Damon Runyon Cancer Research Foundation. C.R.B. is supported by the Howard Hughes Medical Institute. E.C.

is supported by the Pew Charitable Trusts, Charles H. Hood and March of Dime Foundations, Charles E. Reed Faculty Initiative and Irwin and Helen Sizer Career Development Professorship. The project is in part supported by R01CA184187, R01CA158073 and CA215067 to S.Z. S.Z. was a Leukemia Lymphomas Society Scholar. J.L.C. was supported by 1F31CA183504. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30CA013696 to Herbert Irving Comprehensive Cancer Center (HICCC) of Columbia University.

Author contributions The Calo and Zha groups independently uncovered an RNA-dependent role for DNA-PK in ribosome biogenesis and protein synthesis. Z.S., J.L.C., Y.Z. and S.Z. conceived the mouse genetics, hematopoietic stem and progenitor cell-related experiments. Y.Z. measured protein translation and nucleoli localization. J.L.C. and W.J. generated the DNA-PKcs mutant mouse models. V.M.E. and B.J.L. contributed to the generation and characterization of the TP53- and KU70-deficient DNA-PKcs mutant mice. G.B. contributed to pathology analyses of the mouse models. R.A.F. and E.C. conceived the RNA, proteomics and in vitro reconstitution experiments of this project with the help of J.L., F.A., P.A. and C.R.B. All authors interpreted results and wrote the manuscript.

Competing interests The authors declare no competing interests.

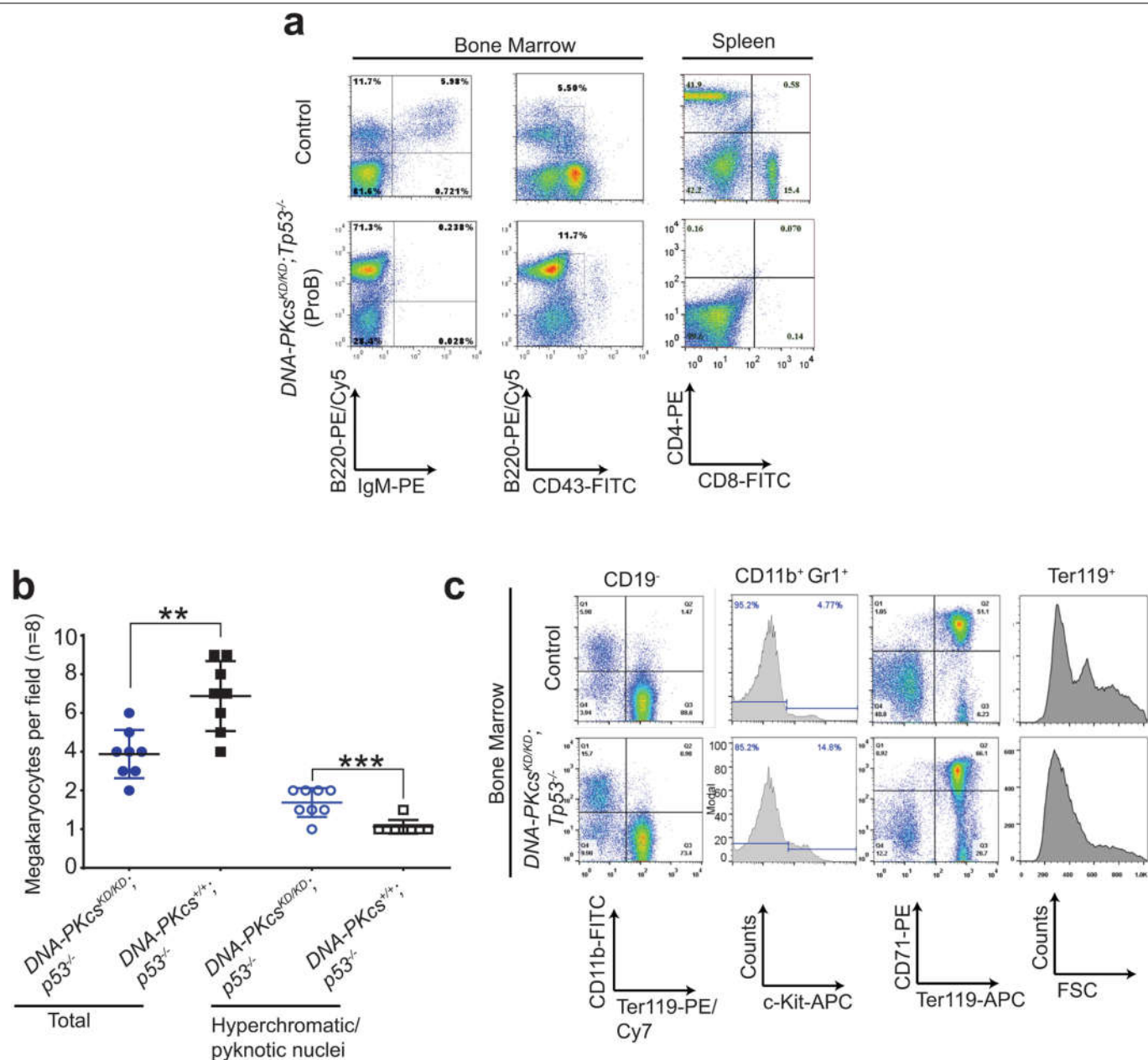
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2041-2>.

Correspondence and requests for materials should be addressed to S.Z. or E.C.

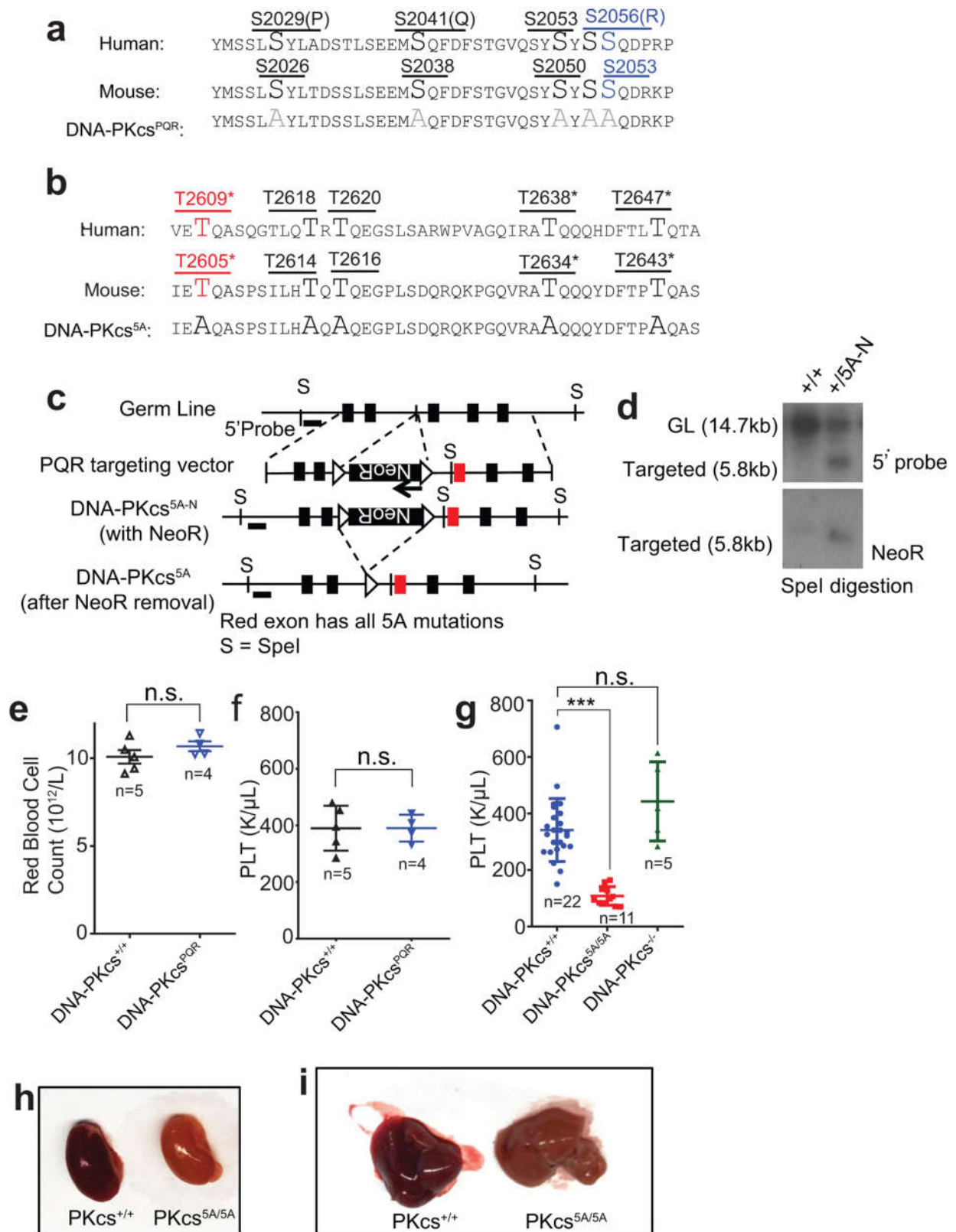
Peer review information *Nature* thanks Alan Warren and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Myeloid proliferative disease with MDS features in $DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice. **a**, Representative fluorescence-activated cell sorting (FACS) analyses of pro-B cell lymphomas from $DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice. **b**, The frequency of megakaryocytes with hyperchromatic or pyknotic nuclei and total megakaryocytes per 400 \times field in the spleen. Eight fields from three independent $DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice were quantified and are presented here. Mean \pm s.e.m.; two-sided unpaired student's *t*-test, ***P* < 0.01, ****P* < 0.001. **c**, Representative FACS analyses of myeloid proliferative disease in $DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice. The CD19-negative bone marrow from

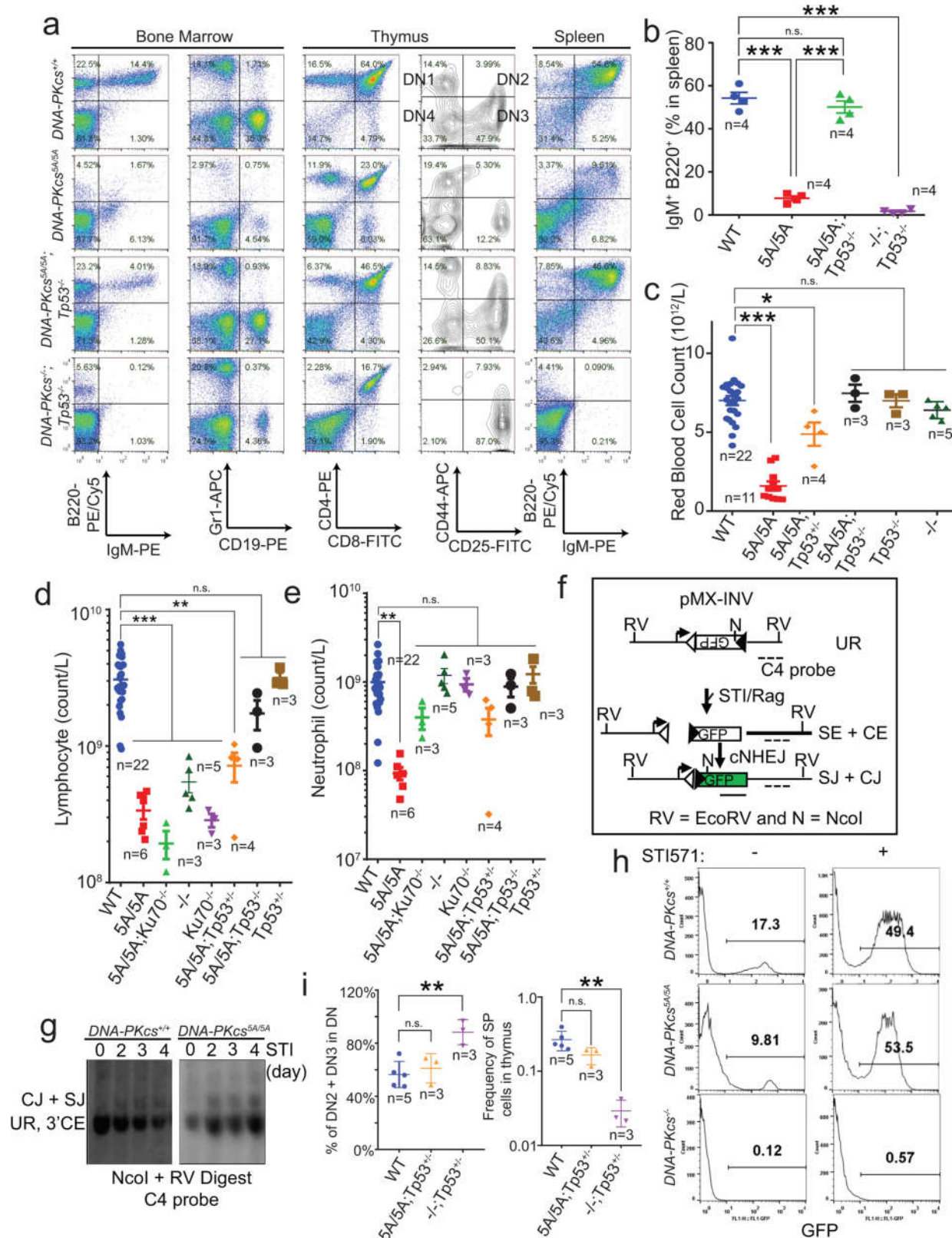
$DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice shows a higher percentage of myeloid cells (CD11b⁺) at the expense of RBCs (TER119⁺). CD19⁺ B cells were excluded from the wild-type bone marrow, as $DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice have no pre-B and mature B cells (CD19⁺) owing to cNHEJ defects. There is also an increase in the percentage of c-KIT⁺ cells among the myeloid cells (CD11b⁺GR1⁺) from $DNA-PKcs^{KD/KD};Tp53^{-/-}$ mice. Mid-sized reticulocytes and RBC progenitors are reduced in the $DNA-PKcs^{KD/KD};Tp53^{-/-}$ bone marrow. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Generation and characterization of the *DNA-PKcs*^{PQR/PQR} and *DNA-PKcs*^{SA/SA} mouse models. **a**, Diagram of the protein sequence changes in the *DNA-PKcs*^{PQR} allele¹⁶. **b, c**, Diagram of the protein sequence changes in the *DNA-PKcs*^{SA} allele and the targeting scheme. The TQ sites are shown in upper case for ease of identification. Asterisks denote sites mutated in the previously published *DNA-PKcs*^{SA} allele¹⁷. The schematic diagram (not to scale) represents the mouse *Prkdc* (*DNA-PKcs*) locus (top), targeting vector (second row), targeted allele (*DNA-PKcs*^{SA}, third row), and the neo-deleted mutant allele (*DNA-PKcs*^{SA}, bottom). The red exon contains the mutations. Boxes, exons; triangles, FRT sites. **d**, Southern blot analyses of

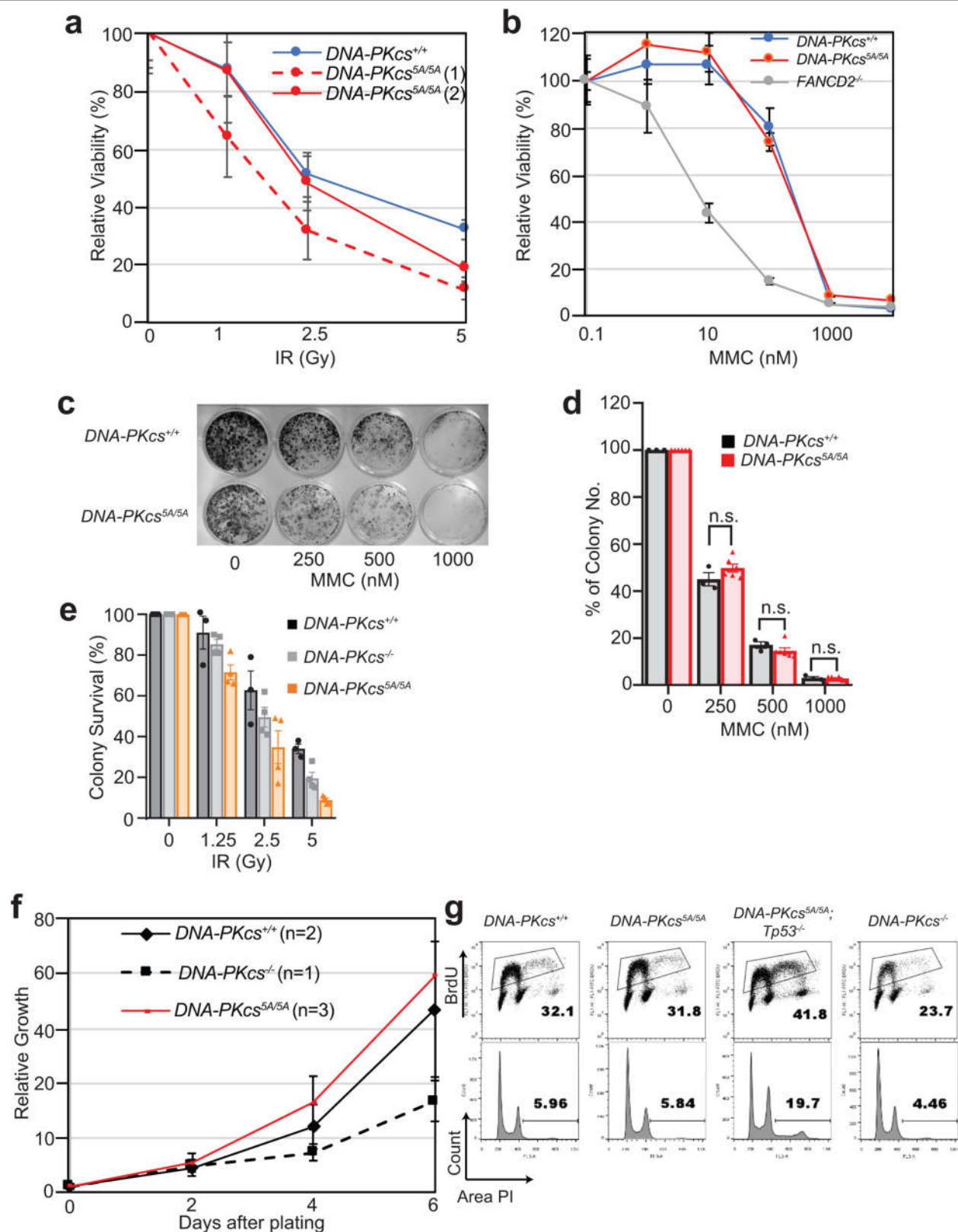
SpeI-digested DNA from *DNA-PKcs*^{+/+} and *DNA-PKcs*^{+/SA} (with NeoR) ES cells. **e, f**, The concentration of RBCs and platelets (PLT) in the peripheral blood of 5-month-old *DNA-PKcs*^{PQR/PQR} mice and *DNA-PKcs*^{+/+} controls. **g**, The platelet concentration in the peripheral blood of 2-week-old *DNA-PKcs*^{+/+}, *DNA-PKcs*^{SA/SA} and *DNA-PKcs*^{-/-} mice. **h, i**, Representative images show pale kidney (**h**) and liver (**i**) from 2-week-old *DNA-PKcs*^{SA/SA} mice, indicating extreme anaemia. **e–g**, Mean ± s.e.m.; two-sided unpaired Student's *t*-test, ****P* < 0.001; n.s., not significant, *P* > 0.05). Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.



Extended Data Fig. 4 | See next page for caption.

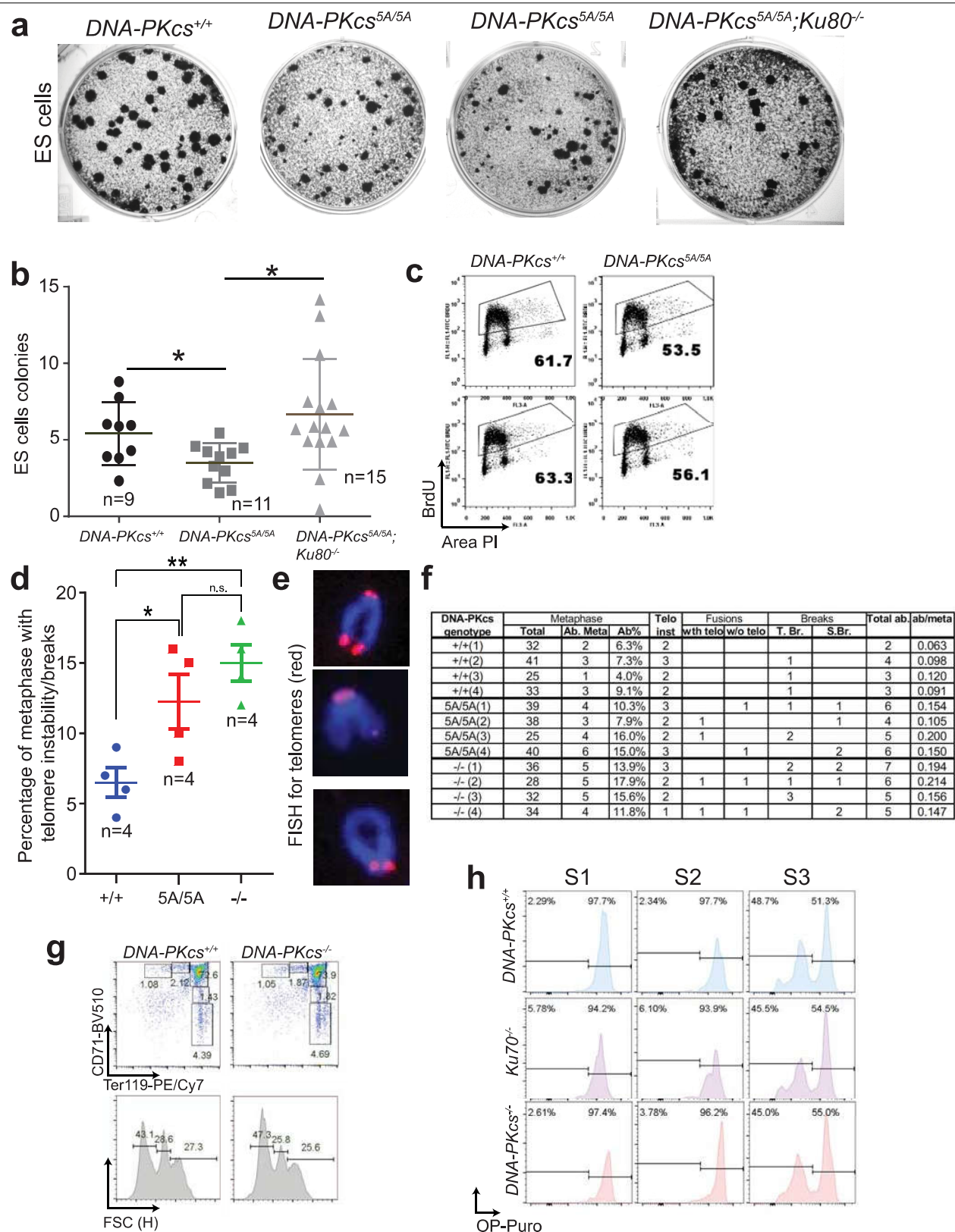
Extended Data Fig. 4 | Normal V(D)J recombination in *DNA-PKcs*^{SA/SA} mice and cells. **a**, Flow cytometry analyses of lymphocyte development in *DNA-PKcs*^{SA/SA}*Tp53*^{-/-} and control mice. **b**, Frequency of B220⁺IgM⁺ naive mature B cells from spleen. **c**, Quantification of peripheral RBCs in *DNA-PKcs*^{SA/SA} mice with and without TP53 deficiency. Note that wild-type, *DNA-PKcs*^{SA/SA} and *DNA-PKcs*^{-/-} data from Fig. 2b are included here for comparison. **d, e**, Absolute counts of peripheral lymphocytes (**d**) and neutrophils (**e**) in *DNA-PKcs*^{SA/SA} mice with or without *Tp53* or *Ku70* deficiency. TP53-deficiency rescued both the lymphocytopenia and the neutropenia in *DNA-PKcs*^{SA/SA} mice. Notably, KU70 deficiency caused severe lymphocytopenia by itself or with the *DNA-PKcs*^{SA/SA} mutation, but rescued the neutropenia in *DNA-PKcs*^{SA/SA} mice. **f–h**, Chromosomal V(D)J recombination measured in v-ABL kinase-transformed B cells. **f**, Diagram of pMX-INV chromosomal V(D)J recombination substrate².

Empty and filled triangles represent recombination signal sequences. RV, EcoRV, N, NcoI. Dashed line marks the C4 probe used for Southern blotting. **g**, Southern blotting reveals the accumulation of normal coding joins (CJ) and signal joins (SJ) and suggests successful V(D)J recombination in *DNA-PKcs*^{SA/SA} B cells. **h**, Successful V(D)J recombination places *Gfp* in the same orientation as the promoter and leads to GFP expression². Plots show the frequency of GFP⁺ cells after 4 days of STI571 (3 μM) treatment. **i**, Quantification of DN2 and DN3 cells out of all double-negative cells (left, reflecting TCRβ V(D)J recombination) and percentage of CD4⁺ or CD8⁺ single-positive (SP) cells (right, reflecting TCRα V(D)J recombination). **b–e, i**, Mean ± s.e.m.; two-sided unpaired Student's *t*-test. ****P* < 0.001, ***P* < 0.01, **P* < 0.05, n.s. *P* > 0.05. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.



Extended Data Fig. 5 | *DNA-PKcs*^{5A/5A} MEFs and ES cells are not hypersensitive to DNA-damaging agents. **a, b**, *DNA-PKcs*^{5A/5A} MEFs are not hypersensitive to ionizing radiation (IR) or mitomycin C (MMC). *Fancd2*^{-/-} cells¹⁹ are included as positive controls. **c–e**, Representative colony formation assays (**c**) and quantification of sensitivity to MMC (**d**) and IR (**e**) in *DNA-PKcs*^{5A/5A} and control ES cells. Mean ± s.e.m. of a representative experiment in triplicate. Two-sided

unpaired Student's *t*-test; *P* > 0.05 (n.s.). **f**, Normal proliferation curve of *DNA-PKcs*^{5A/5A} MEFs. **g**, *DNA-PKcs*^{5A/5A} MEFs have normal S phase frequency. The accumulation of aneuploid cells is unique to *Tp53*^{-/-} cells. **a, d–f**, Mean ± s.e.m. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.



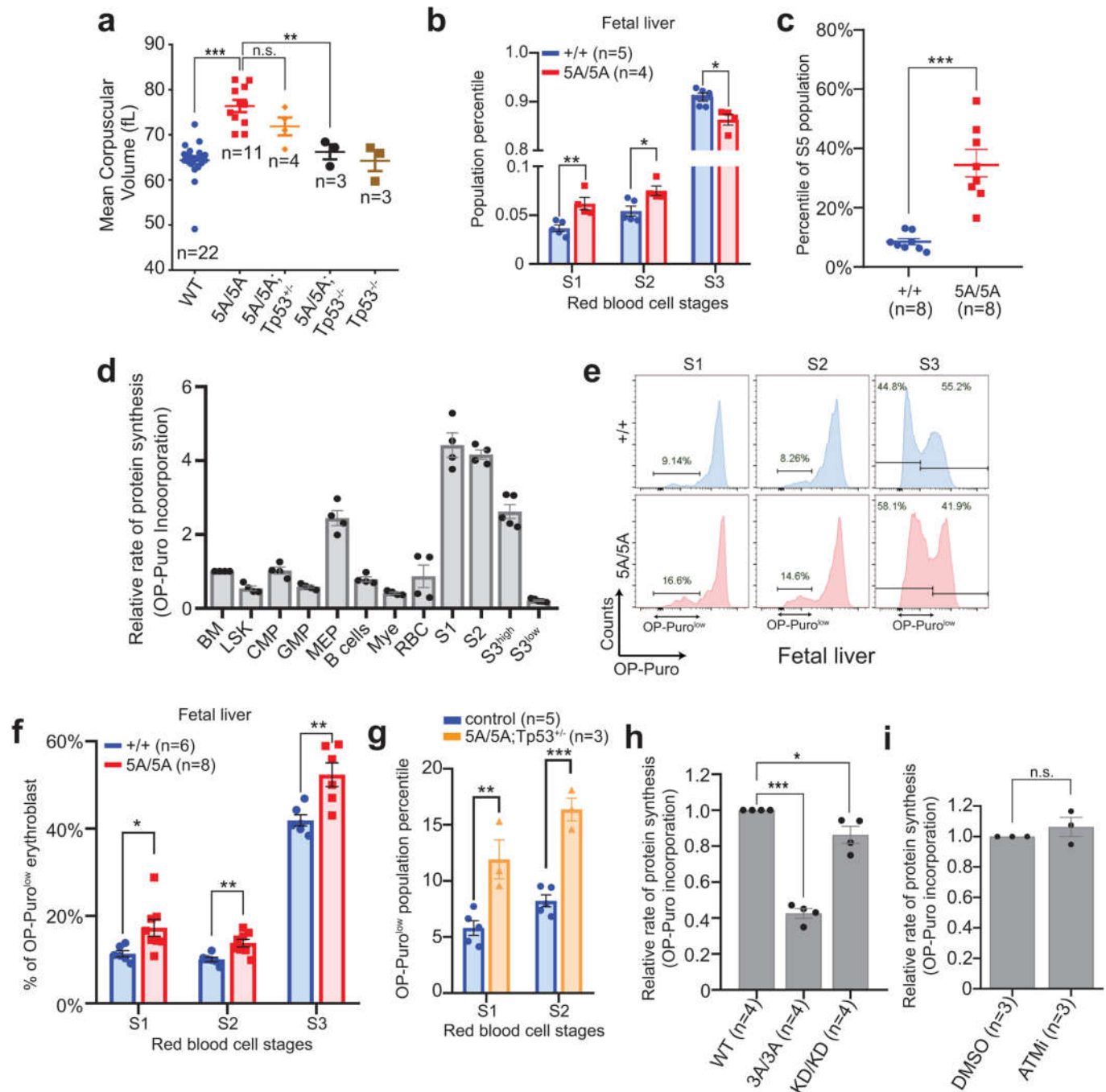
Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | *DNA-PKcs*^{SA/SA} ES cells have KU-dependent proliferation defects that cannot be fully explained by cNHEJ or telomere defects.

a, Representative colony formation assay for *DNA-PKcs*^{SA/SA} ES cells. Note the frequent accumulation of small colonies in *DNA-PKcs*^{SA/SA} cells, which is rescued by deletion of KU80. **b**, Quantification of colony size (A.U., arbitrary units) from *DNA-PKcs*^{SA/SA} and control ES cells. Two independently derived ES cells were assayed. **c**, Cell cycle analyses of two independently derived *DNA-PKcs*^{SA/SA} and control ES cell lines. The percentage of BrdU⁺ cells is shown. **d**, Frequency of metaphase with telomere abnormalities (see below for definition). **e**, Representative telomere fluorescence in situ hybridization (FISH) images of a normal mouse chromosome (top, with four telomere dots), a chromosome with a chromatid break (middle, showing loss of one telomere signal among the two sister chromatids), or a chromatid fusion without telomere signal (bottom). **f**, Quantitative analyses of telomere instability and chromosomal breaks in metaphase. Telomere FISH analyses of MEFs were performed with the telomere-specific PNA probe as previously described². Normal mouse chromosomes have four discrete telomere signals (**e**, top).

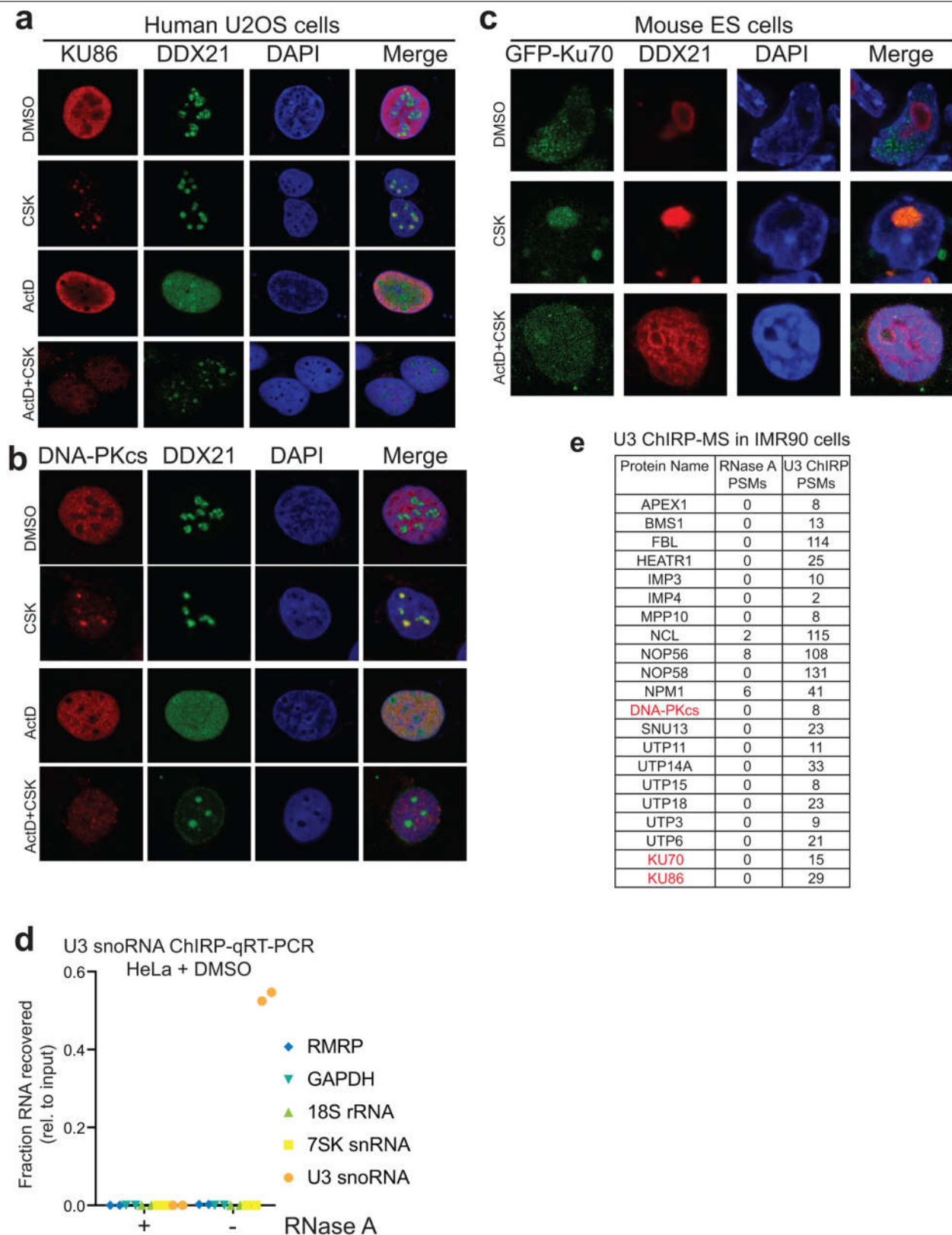
Telomere instability or breaks considered include: i) telomere instability (indicated by more than one telomere signal per chromatid), ii) telomere/chromosome fusion (**e**, bottom; with telomere at the fusion junction (telomere fusion) or without telomere signal at the fusion junction (non-telomere fusion)), iii) chromosome breaks (S.B.; loss of both telomere signals on the paired sister chromatids) and iv) chromatid breaks (T.B.; loss of one of the two chromatids (**e**, middle)). The number of metaphases with at least one telomere instability, break or fusion is shown in **d** as a percentage of metaphases with abnormalities. Data derived from four independent MEF lines of each genotype. **g**, Representative flow cytometry analyses of erythroblasts from age-matched (2 weeks) *DNA-PKcs*^{+/+} and *DNA-PKcs*^{-/-} mice. **h**, Representative protein translation analyses of S1 (CD71⁺Ter119⁻), S2 (CD71⁺Ter119^{mid}) and S3 erythroblasts (CD71⁺Ter119^{high}) from 2-week-old *DNA-PKcs*^{+/+}, *DNA-PKcs*^{-/-} and *Ku70*^{-/-} mice. Quantification is shown in Fig. 2i. **b**, **d**, Mean ± s.e.m.; two-sided unpaired Student's *t*-test, ***P* < 0.01, **P* < 0.05, n.s. *P* > 0.05. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.



Extended Data Fig. 7 | Global translation defects in *DNA-PKcs*^{5A/5A}

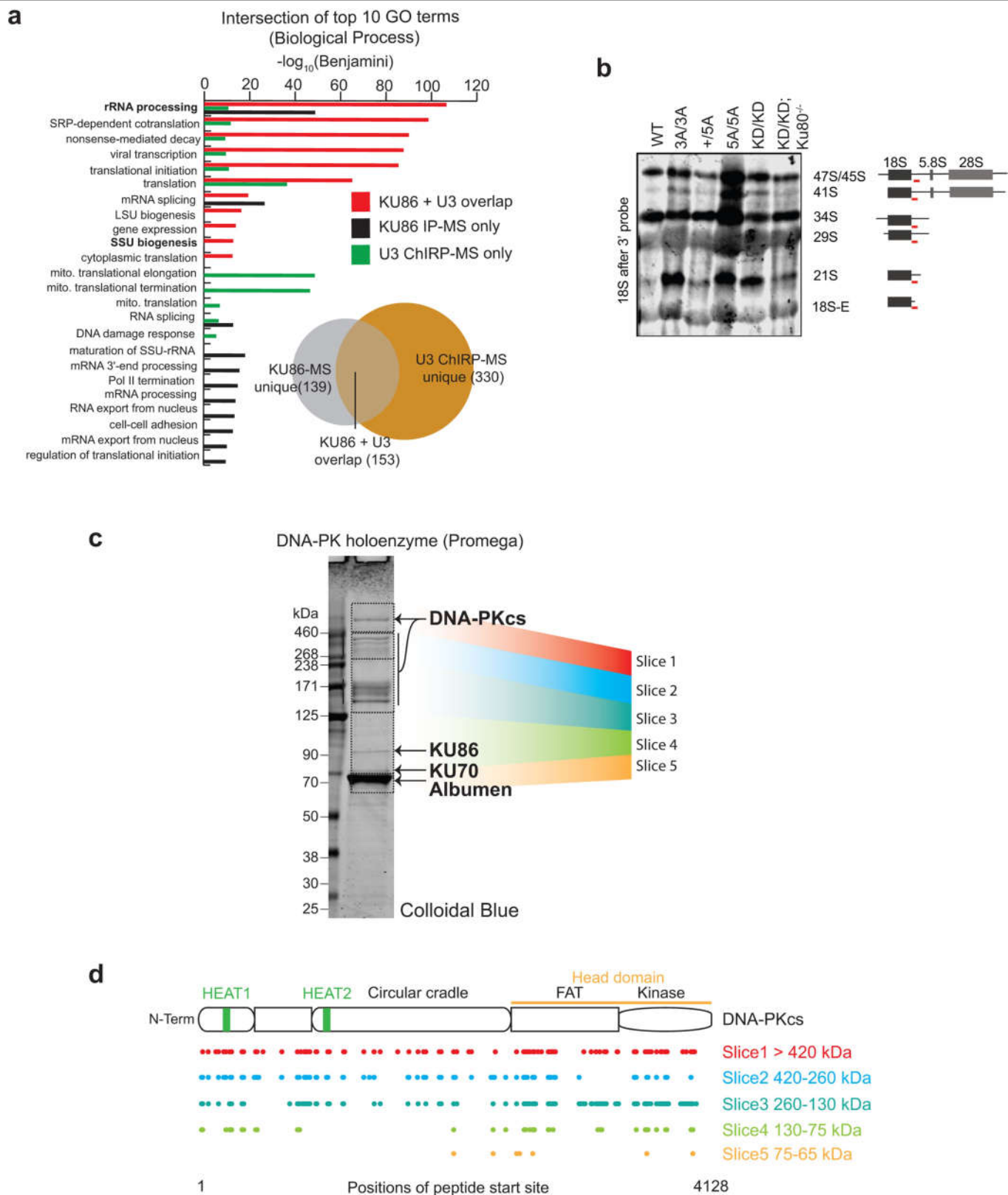
haematopoietic cells. a, MCV of RBCs from 2-week-old *DNA-PKcs*^{5A/5A} mice with or without TP53 deficiency. Wild-type and 5A/5A data from Fig. 2e included for comparison. **b**, Relative percentage of S1, S2 and S3 cells among all erythroblasts in fetal liver from *DNA-PKcs*^{5A/5A} or *DNA-PKcs*^{5A/5A};Tp53^{-/-} E14.5 embryos. S4 and S5 populations have not yet evolved significantly in the E14.5 fetal liver. **c**, The relative frequencies of S5, the most mature erythroblast, in P14 bone marrow from *DNA-PKcs*^{5A/5A} or *DNA-PKcs*^{5A/5A};Tp53^{-/-} mice. **d**, Relative OP-puro levels in P14 wild-type bone marrow haematopoietic cells. The average of total bone marrow was set to 1. The erythroblasts (S1, S2 and S3^{high}) clearly have the highest OP-Puro levels. As shown in Fig. 2h, S3 erythroblasts have an OP-puro^{high} and an OP-puro^{low} population. The mean OP-puro levels of both populations are included. **e**, Representative global protein translation in S1, S2 and S3

erythroblasts from E14.5 *DNA-PKcs*^{5A/5A} and *DNA-PKcs*^{5A/5A};Tp53^{-/-} fetal liver. **f**, Frequency of OP-Puro^{low} among S1, S2, and S3 erythroblasts in E14.5 fetal liver. OP-Puro labelling of fetal liver cells was performed for 30 min. **g**, Quantification of the frequency of OP-Puro^{low} among S1 and S2 erythroblasts from 2-week-old *DNA-PKcs*^{5A/5A};Tp53^{-/-} mice. **h**, Relative OP-puro levels (normalized to the levels in untreated *DNA-PKcs*^{5A/5A} B cells) of *DNA-PKcs*^{5A/5A};Tp53^{-/-} and *DNA-PKcs*^{5A/5A};3A/3A B cells. Unpaired Student's *t*-test, ****P* < 0.001, ***P* < 0.01, **P* < 0.05. **i**, Relative OP-puro levels of *DNA-PKcs*^{5A/5A} B cells treated with ATM kinase inhibitor (KUS5933, 15 μM for 17 h). Data represent three independent biological experiments (*P* = 0.38 for unpaired Student's *t*-test). **a–d**, **f–i**, Two-sided unpaired Student's *t*-test, ****P* < 0.001, ***P* < 0.01, **P* < 0.05, n.s. *P* > 0.05. All graphs show mean ± s.e.m. Exact *P* values and defined sample sizes (*n*) are provided in Supplementary Data 1.



Extended Data Fig. 8 | rRNA synthesis-dependent localization of DNA-PK in nucleoli of human and mouse cells, and U3 ChIRP-MS of SSU processome. **a, b**, Immunofluorescence staining of endogenous KU86 (**a**) and DNA-PKcs (**b**) in U2OS cells. DDX21 RNA helicase is used as a positive control for nucleoli. The CSK buffer contains Triton X-100 for pre-extraction before fixation (see Methods). When indicated, the cells were treated with 50 nM ActD for 1 h before pre-extraction, fixation and staining. **c**, Localization of ectopically expressed GFP-tagged KU70 in mouse ES cells. **a–c**, $n=3$ biologically independent experiments. **d**, U3 ChIRP-qRT-PCR analysis from HeLa cells. Enrichment levels, relative to input samples, of the U3, 7SK, 18S, and RMRP RNAs were assessed from experimental (–RNase A) or control (+RNase A) ChIRP samples. Data are from two independent biological replicates. **e**, DNA-PK was also recovered from U3 ChIRP-MS in IMR90 cells. Peptide spectral match (PSM) counts for control (RNase A) and experimental (U3) samples are shown. $n=2$ biological replicates.

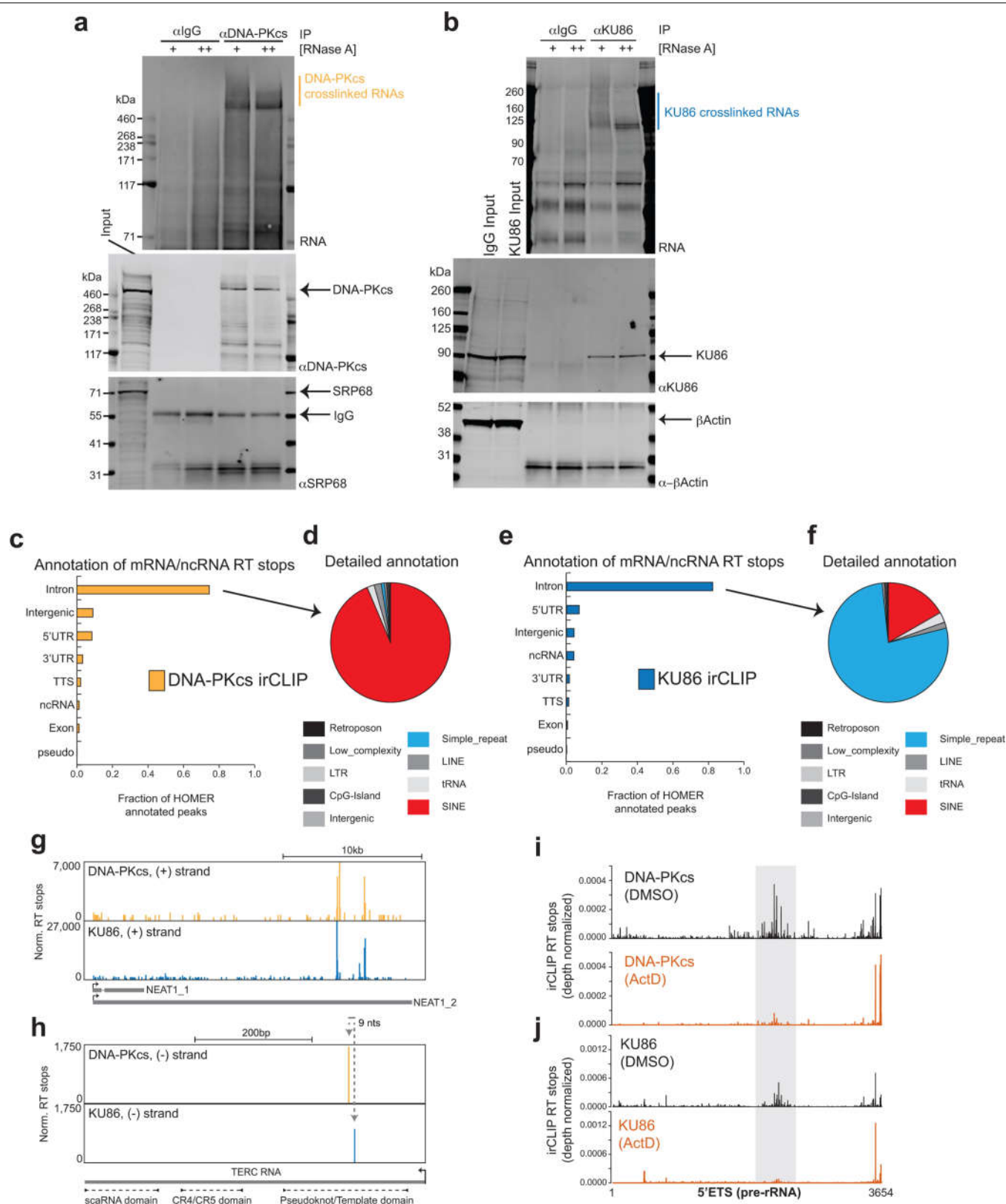
independent experiments. **d**, U3 ChIRP-qRT-PCR analysis from HeLa cells. Enrichment levels, relative to input samples, of the U3, 7SK, 18S, and RMRP RNAs were assessed from experimental (–RNase A) or control (+RNase A) ChIRP samples. Data are from two independent biological replicates. **e**, DNA-PK was also recovered from U3 ChIRP-MS in IMR90 cells. Peptide spectral match (PSM) counts for control (RNase A) and experimental (U3) samples are shown. $n=2$ biological replicates.



Extended Data Fig. 9 | See next page for caption.

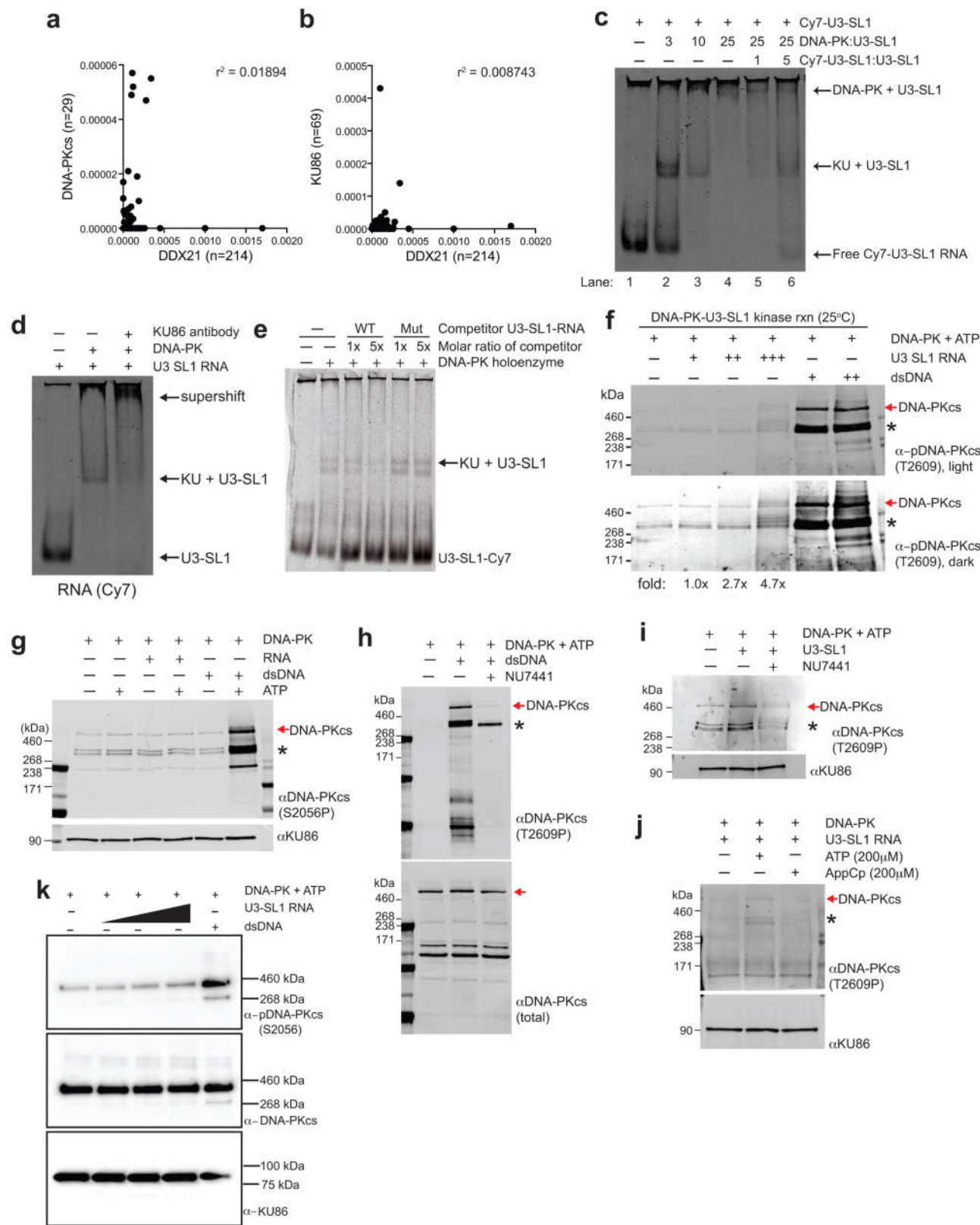
Extended Data Fig. 9 | KU associates with SSU processome components. **a**, KU86 IP-MS and U3 ChIRP-MS overlap. Zhou et al.³³ ($n=2$ independent MS assays; for each assay, $n=2$ technical replicates) identified 292 proteins enriched with KU86 protein. U3 ChIRP-MS identified 483 proteins enriched more than twofold with the U3 snoRNA (Supplementary Table 1). These identified factors were intersected, resulting in 153 proteins in common between the two affinity purification strategies, which is highly significant (hypergeometric $P < 6.650 \times 10^{-166}$). To understand what types of protein were enriched only with KU86, only with U3, or together with both factors, we isolated the enrichment values ($-\log_{10}(\text{Benjamini})$) for GO biological process terms for each of these sets and compared them. Factors commonly bound were biased for rRNA processing, ribosomal terms and SSU biogenesis. U3-specific factors had additional enrichment in these categories and KU86 had a set of unique terms that were not well represented in U3. **b**, Independent repeat of northern blot analyses of 18S rRNA maturation in v-ABL kinase-transformed B cells from noted genotypes. The probe covers the sequence just after the 18S

rRNA (red line). This experiment was repeated independently four times. Another repeat is shown in Fig. 3c. **c**, **d**, MS characterization of commercial DNA-PK holoenzyme (Promega) used in EMSA and kinase reactions. A detailed description of proteins and their quality in this mixture has not been published. We subjected the DNA-PK enzyme mix as provided to SDS-PAGE separation followed by LC-MS identification of proteins from mass ranges between 65 and 600 kDa. KU70 and KU86 were clearly present in the gel and via MS. For masses above 130 kDa, DNA-PKcs was the major protein identified. For each of the five slices analysed (coloured regions) we tabulated the starting positions of peptides matching the DNA-PKcs polypeptide and mapped them to the position within DNA-PKcs. As expected, in the highest-molecular-weight slices, we identified peptides across the majority of the length of DNA-PKcs. DNA-PKcs peptides were present in lower slices, but poorer overall coverage was evident, suggesting that these are degradation products. As this was a confirmatory experiment of a validated and commercially available product (see Methods), it was conducted only once.



Extended Data Fig. 10 | DNA-PKcs and KU86 bind nuclear and nucleolar RNAs in living cells. **a, b**, irCLIP of DNA-PKcs (**a**) and KU86 (**b**) in HeLa cells. RNA (IR800) and protein blots show specific enrichment of RNA UV-crosslinked to DNA-PKcs and KU86. The increased signal in the RNaseA-treated samples corresponds to the accumulation of RNaseA-protected fragments that are directly associated with either KU86 or DNA-PKcs. Results shown are representative of three biologically independent experiments. Vertical line marks the RNA extracted for sequencing. **c, e**, Fractional distribution of DNA-

PKcs (**c**) and KU86 (**e**) irCLIP RT stops on major genomic features annotated with HOMER. **d, f**, Detailed HOMER annotation of DNA-PKcs (**d**) and KU86 (**f**) RT stops categorized as intronic and intergenic from **c** and **e**. **g, h**, Genome browser tracks of DNA-PKcs and KU86 irCLIP on the *Neat1* (**g**) and *Terc* (**h**) RNA. RT stop values are normalized to 10^7 mapped reads. **i, j**, Histograms of DNA-PKcs (**i**) and KU86 (**j**) irCLIP RT stops mapping to the 5'ETS, normalized for total reads mapped per experiment. Data from DMSO- (top) or ActD-treated (bottom) HeLa cells are shown. Grey highlighted region is as in Fig. 4a.



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | DNA-PK interacts with structured RNAs that can activate its kinase. **a, b,** Correlation analysis of total RT stops mapping to non-repeat snoRNA transcripts from DDX21³⁵ compared to DNA-PKcs (**a**) or KU86 (**b**) irCLIP experiments from DMSO-treated HeLa cells. Correlation analysis was performed using Pearson's correlation coefficient. *n* denotes number of snoRNA transcripts bound by each protein. **c,** EMSA of purified human DNA-PK and in vitro transcribed U3-SL1. Lane 1 contains only Cy7-labelled U3-SL1. Lanes 2–4 show that KU assembles with U3-SL1 at a 1:3 molar ratio, while DNA-PK holoenzyme assembly occurs at a 1:25 molar ratio. Lanes 5 and 6 show that unlabelled U3-SL1 RNA competes away bound labelled U3-SL1 in a dose-dependent manner. **d,** Supershift EMSA of DNA-PK and U3-SL1 RNA with KU86 antibody. The addition of anti-KU86 confirms the identity of the KU–U3-SL1 band and also shifts up the complex to higher molecular weights. **e,** A structural mutant of U3-SL1 was generated by introducing point mutations predicted to disrupt the stem-loop structure. This mutant was unable to compete away wild-type U3-SL1 for binding to the KU complex, while unlabelled wild-type U3-SL1 competed efficiently. **f,** DNA-PK in vitro kinase phosphorylation assay in the presence of increasing amounts of U3-SL1 or DNA. Western blot was performed

with an antibody recognizing DNA-PKcs phosphorylated at the T2609 cluster. Asterisks denote cross-reactive fragments that probably include phosphorylated DNA-PKcs fragments, on the basis of MS analyses of the DNA-PK complex (Extended Data Fig. 9c, d). **g,** As in **f**, but using an antibody recognizing DNA-PKcs phosphorylated at the S2056 cluster. **h–j,** As in **f** with the following changes. **h,** dsDNA was used to activate DNA-PK, NU7441 was included to inhibit specific DNA-PK activity, and western blot analysis monitored the total DNA-PK (total DNA-PKcs) or phosphorylated DNA-PK (DNA-PKcs phoT2609). **i,** U3-SL1 RNA was used to activate DNA-PK in the absence or presence of the DNA-PK inhibitor NU7441. **j,** U3-SL1 RNA was used to activate DNA-PK, hydrolysable (ATP) or non-hydrolysable (AppCp) ATP was provided, and western blot analysis monitored KU86 (loading control) or phosphorylated DNA-PK (phoT2609). **k,** Baculovirus-purified human DNA-PK in vitro kinase phosphorylation assay in the presence of increasing amounts of U3-SL1 or DNA. Western blot was performed with antibodies recognizing DNA-PKcs phosphorylated at the S2056 cluster (top), total DNA-PK (middle) and KU86 (bottom). All EMSA and western blots presented here are representative of three biologically independent experiments.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

irCLIP data were collected in biological duplicate. Sequencing data were generated on Illumina NextSeq 500 instruments. Flow cytometry data were collected either using BD FACSDiva software (V. 8.0.1) or with BD CellQuest Pro (V. 6.0).

Data analysis

irCLIP data were analyzed using a custom script available at: <https://github.com/ChangLab/FAST-iCLIP/tree/lite>. For Images analyses we used MATLAB 2016b (MathWorks). Flow cytometry data were analyzed using FlowJo v. 10. And mouse phenotype data were analyzed using Graphpad Prism V. 8.0. for p value with statistic method specified in figure legends. GO term analysis of the U3 ChIRP-MS and Ku86 MS was perform using the DAVID tool (<https://david.ncifcrf.gov/>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

irCLIP data is available via GEO under the accession number GSE109026. All uncropped blots are provided in Supplementary Figure 1. Data underlying the graphical representations used in the figures, including all experiments presenting data from animal models, are provided as a Source Data File. Exact (p) values and defined sample size (n) are provided in Supplementary Data 1.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For each experimental type, an appropriate number of data points or experiments were collected. For animal or cell line based experiments, at least 3 independently derived samples were measured. For example, deep sequencing experiments were performed in duplicate, each mice genotype data were collected on more than 3 individuals from more than 2 independent crosses.
Data exclusions	No data were excluded from the analysis.
Replication	As no data were excluded, all values from each experiment are presented and the variation can be seen in the figures. Statistical analysis as described throughout the manuscript provided the ability to confidently assess differences between different experimental conditions. More information and details can be found on Supplemental Data 1. All mouse phenotype analyses were done multiple times independently with more than 3 mice from at least 2 independent crosses. Only results that can be successfully repeated were included in the manuscript.
Randomization	Randomization was not performed, however biological replicate experiments were performed on different days. For the animal and mouse cell line based experiments, the samples are allocated based on genotypes (confirmed and validated via PCR).
Blinding	No blinding was used during the experimentation. The mutant mice were visibly smaller, genotypes have to be determined ahead of the experiments to plan complicated analyses.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

[Human Antibodies for Westerns (dilution 1:2000): pDNA-PKcs S2056 (Abcam, ab18192, lot. GR296008-1) and T2609 clone 10B1 (ab18356, lot. GR309667-22), DNA-PKcs (Novus Biologicals, NBP2-33995, lot. R32563), DNA-PKcs (Bethyl, A303-967A), Ku80 (Bethyl, A302-627A and Novus, NBP2-373583, lot. 140928) and p53 clone 1C12 (Cell Signaling, 2524S, lot. 13)]. Rabbit IgG (ThermoFisher Scientific, 02-6102, lot. Dilution 1:10,000). Secondary antibodies include Alexa-Fluor 488, 568, and 647 (Life Technologies). Antibodies for FACS analysis: biotin anti-mouse CD4 (cat. 100404, clone GK1.5, lot. B245033), biotin anti-mouse

CD8a(cat. 100704, clone 5.3-6.7, lot. B257790), biotin anti-mouse CD3e(cat. 100304, clone 145-2C11, lot. B268248), biotin anti-mouse B220(cat. 103204, clone RA3-6B2, lot. B268525), biotin anti-mouse Ly6G/Ly6C (Gr-1) (cat. 108404, clone RB6-8C5, lot. B236916), biotin anti-mouse TER119(cat. 116204, clone TER-119, lot. B256623), FITC anti-mouse CD8a (cat. 100706, clone 53-6.7, lot. B259672), PE/Cy7 anti-mouse CD16/32(cat. 101318, clone 93, lot. B249751), PE/Cy7 anti-mouse TER119(cat. 116222, clone TER-119, lot. B251241), APC anti-mouse Ly6G/Ly6C(Gr-1) (cat. 108412, clone RB6-8C5), APC anti-mouse TER119(cat. 116212, clone TER-119, lot. B246740), APC anti-mouse CD117(c-Kit) (cat. 135108, clone ACK2, lot. B255217), APC/Cy7 Streptavidin (cat. 405208, lot. B258702), Brilliant Violet 510 anti-mouse CD41 (cat. 133923, clone MWRReg30, lot. B258166), Brilliant Violet 510 anti-mouse CD71 (cat. 113823, clone RI7217, lot. B301604) and Pacific Blue anti-mouse Ly-6A/E(Sca-1) (cat. 108120, clone D7, lot. B258720) were purchased from BioLegend. APC anti-mouse CD44(cat. 17-0441-82, clone IM7, lot. 4312474), and PE-Cyanines5 anti-Hu/Mo CD45R(B220)(cat. 15-0452-83, clone RA3-6B2, lot. 2052744) were purchased from eBioScience. FITC rat anti-mouse CD25(cat. 1595-02S, lot. 17103-P624AF), Goat anti-mouse IgM (cat. 1020-09, lot. K2915-5357) were purchased from Southern Biotech). PE rat anti-mouse CD4(cat. 557308, clone GK1.5, lot. 8093955), PE rat anti-mouse CD19 (cat. 557399, clone 1D3, lot. 7213714) and biotin rat anti-mouse CD5(cat. 553018, clone 53-7.3, lot 8115630), FITC anti-mouse CD43(cat. 553270, clone S7, lot. 6110908), and Alexa Fluor 700 rat anti-mouse CD34(cat. 560518, clone RAM34, lot. 8116836) were purchased from BD Pharmingen. All FACS antibodies were used at 1:200 dilution. Antibodies used for Immunofluorescence: Mouse anti-Hu/Mo Ku80(111) monoclonal antibody (MA5-12933) was purchased from ThermoFisher and used at 1:100 dilution. Rabbit anti DDX21 polychonal antibody (NB100-1718) was purchased from Novus and used at 1:500 dilution. Mouse anti-DNA-PKcs Ab-4 (Cocktail) monoclonal antibody (MS423P) was purchased from ThermoFisher and used at 1:100 dilution. Alexa Fluor 488 conjugated goat anti-Rabbit secondary antibody (cat. A-11008, lot. 1937184), Cyanine3 conjugated goat anti-mouse secondary antibody (cat. A10521, lot. 1126597), and Alexa Fluor 594 conjugated goat anti-Rabbit secondary antibody (cat. A-11012, lot. 1745478) were purchased from Invitrogen and used at 1:500 dilution.

Validation

All antibodies are extensively used and well accepted commercial antibodies and validated in several published manuscript. Antibodies were further validated in our lab by western blotting appropriate cell lysates and assessing the resulting western blot for signal where the band of the expected molecular weight of the target protein would constitute at least 80% of the total signal of that lane. In some cases we performed IP mass spec to further validate the specificity of the antibody. FACS antibodies were validated by single staining on positive population from WT mouse, and Immunofluorescence antibodies were validated by either comparative staining of WT cell line to corresponding knock out cell line, or by inhibitor treatment, for intensity and position of positive signal. See methods for details. Additionally, our in vitro kinase assay with in vitro purified humans proteins further validate the specificity of our DNA-PK, Ku86 and p53 antibodies.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

HeLa and IMR90 were purchased from ATCC and Sigma, respectively. U2OS cells were purchased from ATCC. Most MEFs, ES cells and v-abl kinase transformed B cells were freshly derived from genetically modified animal models in the Zha lab, unless noted below. Standard protocols (can be found in references) were used to derived those cell lines. Their genotypes were confirmed and validated via PCR. More than 2, often more than 3 independently derived cell lines of each genotypes were used for each measurement. A few control cells were generously provided by colleagues. Specifically, the DNA-PKcs^{-/-} Abelson lines and the KU70^{-/-} ES cells were originally derived in the laboratory of Dr. Frederick Alt at the Children's Hospital at Boston. The DNA-PKcs3A/3A Abelson lines were generously provided by Dr. Barry Sleckman then at Cornell University and now at the University of Alabama. Fand2^{-/-} MEFs were generously provided by Dr. Agata Smogorzewska at the Rockefeller University. The genotype of all these cell lines have been independently verified by PCR in the Zha lab to the best of our ability.

Authentication

Commercial available cell lines were authenticated by the vendor, we have not re-authenticated the cells since we bought them. Investigator derived cell lines were validated by PCR genotyping to the best of our ability.

Mycoplasma contamination

All cell lines are mycoplasma free. We routinely test for mycoplasma contamination.

Commonly misidentified lines (See [ICLAC](#) register)

None of the cell lines used in this study are in the database of commonly misidentified cell lines.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

This study includes experiments conducted on house mice (*Mus musculus*) of either pure 129S1/SvImJ strain, or 129S1/SvImJ and C57BL/6J mixed strain, and of both male and female. No gender bias were noted. All analysis, unless otherwise noted (e.g., tumor and survival cohorts), was conducted on mice of about 2 weeks old, or E14.5 embryos.

Wild animals

The study did not involve wild animals

Field-collected samples

The study did not involve samples collected from the field

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Bone marrow cells were collected by flushing femora and tibia stems of both lower limbs, spleen cells were collected by mincing the whole spleen with frosted slides, treating with ACK buffer to get rid of red blood cells unless performing peripheral whole blood compartment studies, and filtered with Falcon 70 micrometer nylon cell strainer. E14.5 fetal liver cells were collected by separating fetal livers from E14.5 feti, and further mincing the livers with frosted slides. All population staining was conducted directly on live cells, and translation studies were conducted by labeling live cells with O-Propargyl-puromycin in RPMI 1640 medium containing 15% FBS for 1 hr, and then samples were fixed with 4% paraformaldehyde, permeabilized with 0.1% saponin before conducting CLICK-IT assay according to manufacture protocol.

Instrument

All FACS data was collected either on LSR II flow cytometry system from BD, or on FACSCalibur flow cytometry system from BD.

Software

FACS data conducted on LSR II was collected using BD FACSDiva software, and FACS data conducted on FACSCalibur was collected using BD CellQuest Pro. All FACS data was further gated and analyzed using Flowjo V10.

Cell population abundance

No cell sorting was conducted in this study.

Gating strategy

All gating were conducted following well accepted gating strategies of corresponding antibodies staining from published studies before. On preliminary FSC/SSC plot, gates were set on gaps between debris, dead cells and live cells. And all live cells were included in further analysis.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Structure of the M2 muscarinic receptor– β -arrestin complex in a lipid nanodisc

<https://doi.org/10.1038/s41586-020-1954-0>

Received: 11 August 2019

Accepted: 7 January 2020

Published online: 16 January 2020

 Check for updates

Dean P. Staus^{1,2,10}, Hongli Hu^{3,8,10}, Michael J. Robertson^{3,4,10}, Alissa L. W. Kleinhenz^{1,2,9}, Laura M. Wingler^{1,2}, William D. Capel¹, Naomi R. Latorraca^{3,5,6}, Robert J. Lefkowitz^{1,2,7}✉ & Georgios Skiniotis^{3,4}✉

After activation by an agonist, G-protein-coupled receptors (GPCRs) recruit β -arrestin, which desensitizes heterotrimeric G-protein signalling and promotes receptor endocytosis¹. Additionally, β -arrestin directly regulates many cell signalling pathways that can induce cellular responses distinct from that of G proteins². In contrast to G proteins, for which there are many high-resolution structures in complex with GPCRs, the molecular mechanisms underlying the interaction of β -arrestin with GPCRs are much less understood. Here we present a cryo-electron microscopy structure of β -arrestin 1 (β arr1) in complex with M2 muscarinic receptor (M2R) reconstituted in lipid nanodiscs. The M2R– β arr1 complex displays a multimodal network of flexible interactions, including binding of the N domain of β arr1 to phosphorylated receptor residues and insertion of the finger loop of β arr1 into the M2R seven-transmembrane bundle, which adopts a conformation similar to that in the M2R–heterotrimeric G_o protein complex³. Moreover, the cryo-electron microscopy map reveals that the C-edge of β arr1 engages the lipid bilayer. Through atomistic simulations and biophysical, biochemical and cellular assays, we show that the C-edge is critical for stable complex formation, β arr1 recruitment, receptor internalization, and desensitization of G-protein activation. Taken together, these data suggest that the cooperative interactions of β -arrestin with both the receptor and the phospholipid bilayer contribute to its functional versatility.

Activation of GPCRs leads to heterotrimeric G-protein-mediated signalling that quickly returns to basal levels¹. This remarkably conserved process of GPCR ‘desensitization’ (Fig. 1a) is mainly orchestrated by two small families of proteins, GPCR kinases and arrestins (reviewed in ref. ⁴). GPCR kinases phosphorylate agonist-bound GPCRs on their carboxyl (C) terminus or intracellular loops (ICLs), leading to the recruitment of arrestin (Fig. 1a). In humans, visual arrestin (arrestin-1) and X-arrestin (arrestin-4) are selectively expressed in the retina, whereas the ubiquitously expressed β -arrestins 1 and 2 (also known as arrestin-2 and arrestin-3, respectively) regulate the hundreds of GPCRs that are found elsewhere. Arrestins comprise juxtaposed N- and C-terminal seven-stranded β -sandwich domains with a central crest of three loops (finger, middle and C-loops)^{5,6}. After the phosphorylated GPCR C terminus engages the N domain of arrestin^{7,8}, conformational changes promote the binding of central crest elements to the receptor seven-transmembrane (7TM) bundle, which sterically blocks the coupling of G proteins⁵. β -arrestins also act as adaptors for

endocytic machinery, thereby increasing receptor internalization⁹. As well as modulating desensitization, β -arrestins potentiate many signalling pathways independently of G proteins². Notably, certain ‘biased’ GPCR agonists preferentially activate G-protein or β -arrestin pathways; this could be exploited therapeutically to obtain more selective drugs^{2,10}.

Numerous high-resolution structures of GPCR–G protein complexes have been obtained, primarily by cryo-electron microscopy (cryo-EM) (reviewed in ref. ¹¹). However, to our knowledge, the only GPCR–arrestin structure reported so far is a crystal structure of rhodopsin fused to a constitutively active mutant of visual arrestin^{8,12}. Therefore, knowledge of the molecular framework of GPCR–arrestin interactions remains limited—especially for the β -arrestins, which modulate the vast majority of GPCRs. Here we report the cryo-EM structure of β arr1 in complex with M2R in high-density lipoprotein (HDL) particles (lipid nanodiscs) that mimic a native membrane. This structure provides new insights into β -arrestin-mediated GPCR desensitization and signalling, and

¹Department of Medicine, Duke University Medical Center, Durham, NC, USA. ²Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC, USA. ³Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. ⁴Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. ⁵Department of Computer Science, Stanford University, Stanford, CA, USA. ⁶Biophysics Program, Stanford University, Stanford, CA, USA. ⁷Department of Biochemistry, Duke University Medical Center, Durham, NC, USA. ⁸Present address: School of Life and Health Sciences, Kobilka Institute of Innovative Drug Discovery, The Chinese University of Hong Kong, Shenzhen, China. ⁹Present address: School of Medicine, University of Michigan, Ann Arbor, MI, USA. ¹⁰These authors contributed equally: Dean P. Staus, Hongli Hu, Michael J. Robertson. ✉e-mail: lefko001@receptor-biol.duke.edu; yiorgo@stanford.edu

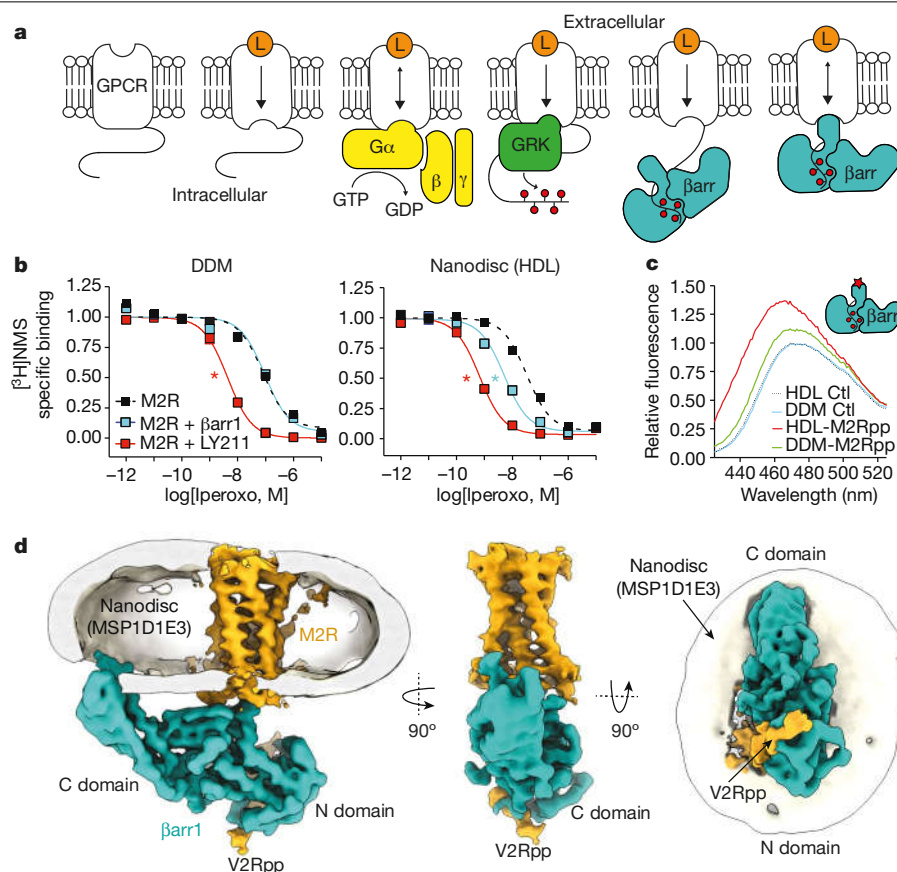


Fig. 1 | Recruitment of β arr1 by M2R in a native lipid environment. **a**, Ligand (L)-induced conformational changes in GPCRs lead to heterotrimeric G-protein activation (GTP hydrolysis) and subsequent GPCR-kinase-mediated receptor phosphorylation. The initial binding of β -arrestin (β arr) to phosphorylated receptors leads to its coupling to the transmembrane (TM) bundle, sterically occluding further G-protein binding. **b**, β arr1 allosterically enhances the affinity of iperoxo to HDL-M2Rpp but not to DDM-M2Rpp, as determined by competition radioligand binding assays. The positive allosteric modulator LY211960 (LY211) enhanced iperoxo affinity regardless of the reconstitution environment. Data are the mean of three independent experiments with error bars representing s.e.m. * indicates significance compared to control (Ctl) (one-way ANOVA). $\log IC_{50}$ values: DDM-Ctl, -7.10 ± 0.09 ; DDM- β arr1,

-7.03 ± 0.06 ; DDM-LY211, -8.36 ± 0.04 ($P < 0.0001$); HDL-Ctl, -7.49 ± 0.08 ; HDL- β arr1, -8.29 ± 0.08 ($P < 0.0007$); HDL-LY211, -9.17 ± 0.05 ($P < 0.0001$). **c**, HDL-M2Rpp, but not DDM-M2Rpp, enhances the fluorescence of bimane in the finger loop of β arr1 (red star, inset). Curves represent the difference in spectra obtained with an antagonist (atropine) and an agonist (iperoxo). Data represent the means of three independent experiments. **d**, Orthogonal views of the cryo-EM density map of the HDL-M2Rpp- β arr1 complex coloured by subunit (orange, M2Rpp; teal, β arr1; grey/white, HDL particle). The orange-coloured density on β arr1 corresponds to the phosphorylated C-terminal peptide (V2Rpp) ligated to the receptor. The nanodisc density, omitted in the middle panel for clarity, was generated earlier in image processing before high-resolution refinement of the M2Rpp- β arr1 complex.

highlights the importance of the membrane environment in these processes.

Structural determination of M2R- β arr1

The stabilization of GPCR- β -arrestin complexes has been historically difficult because of both the need for receptor phosphorylation and the low strength of the β -arrestin-7TM interaction. To reconstitute a functional complex we chose M2R, a family A GPCR that regulates cardiac function¹³, because β arr1 has a relatively strong interaction with its 7TM core¹⁴. To ensure homogeneous phosphorylation, we used sortase to enzymatically ligate a synthetic phosphopeptide (pp) derived from the C terminus of vasopressin 2 receptor (V2R) onto the C terminus of M2R, forming M2Rpp¹⁴ (Extended Data Fig. 1a). Although wild-type M2R lacks a C terminus with phosphorylatable residues and is phosphorylated within ICL3, β arr1 enhances the affinity of agonists for M2Rpp to the same extent as for natively phosphorylated M2R^{14,15}. To enhance the stability of β arr1 we generated a minimal cysteine variant truncated at residue 393 (β arr1(MC-393)), which displayed the same allosteric coupling to M2R as did the wild type (Extended Data Fig. 1b). Hereafter, β arr1 refers to this variant unless otherwise specified.

Initially we characterized the ability of β arr1 to bind M2Rpp in *n*-dodecyl- β -D-maltoside (DDM-M2Rpp) detergent. β arr1 and the antibody fragment Fab30, which stabilizes the β arr1-V2Rpp interaction, co-immunoprecipitate with the receptor only when it is phosphorylated (Extended Data Fig. 1c). Because β arr1 binding to the 7TM core of the receptor enhances agonist binding^{14,15}, we monitored this interaction using radioligand competition binding assays to measure agonist affinity. The affinity of the agonist iperoxo for DDM-M2Rpp is increased by the positive allosteric modulator LY2119620 but not by β arr1 (Fig. 1b). By contrast, when M2Rpp is reconstituted into nanodiscs (HDL-M2Rpp)—which provide a more physiological-like lipid bilayer—both LY2119620 and β arr1 increase the affinity of iperoxo for the receptor (Fig. 1b). The importance of the lipid bilayer in β arr1-receptor binding was further tested by labelling the β arr1 finger loop—which inserts into the 7TM core—with the probe monobromobimane (mBr), the fluorescence of which increases in hydrophobic environments. For these experiments we used a labelled mutant with a valine-to-cysteine substitution at position 70 (V70C); this variant was shown to still couple to the 7TM core of M2Rpp by radioligand binding experiments (Extended Data Fig. 1b). We observed an agonist-dependent increase in the fluorescence of bimane with HDL-M2Rpp but not with DDM-M2Rpp (Fig. 1c, Extended

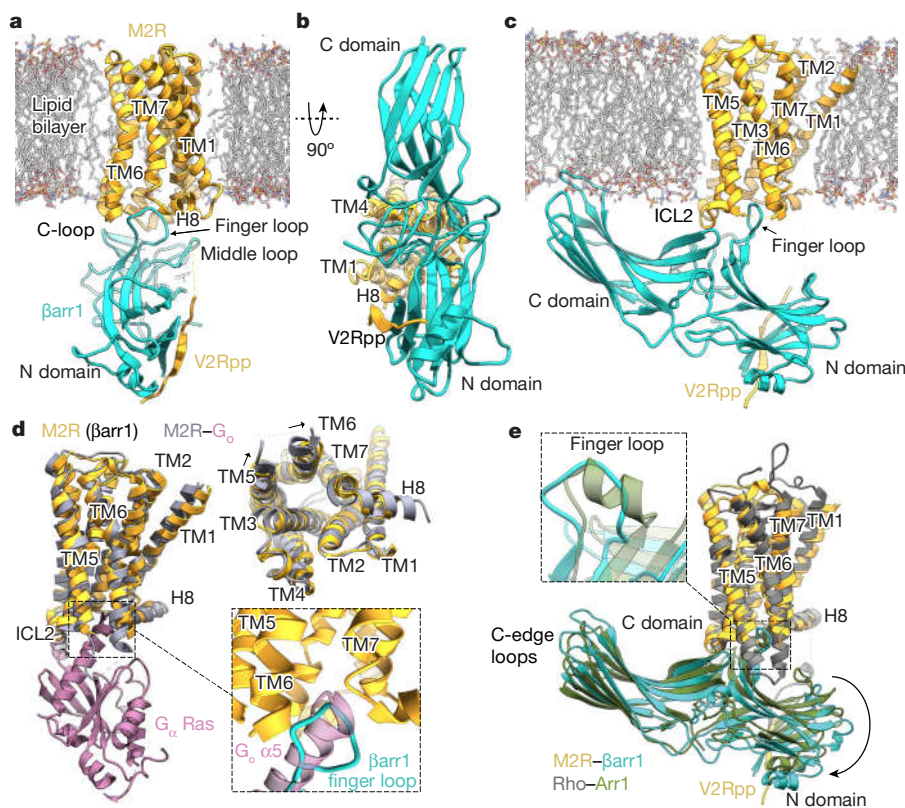


Fig. 2 | Structure of the M2Rpp- β arr1 complex. **a–c**, Orthogonal views of the M2Rpp- β arr1 structure coloured by subunit (orange, M2Rpp; teal, β arr1; grey sticks, model lipid bilayer). **d**, Superposed structures of M2Rpp- β arr1 and M2R- G_{α} complexes (M2R, grey; G_{α} Ras domain, pink; PDB: 6OIK) aligned by M2R. G_{α} $\beta\gamma$ subunits are omitted for clarity. The top right image shows the cytoplasmic view of M2R transmembrane alignment in the absence of transducers. The bottom right inset shows the similar insertion depth of the finger loop of β arr1 and the $\alpha 5$ -helix of G_{α} into the TM bundle of M2R. **e**, Superposition of M2Rpp- β arr1 and Rho-Arr1 structures (rhodopsin, grey; arrestin1, green; PDB: 4ZWI). The inset shows an enlarged view of the finger loops of β arr1 (M2Rpp) and visual arrestin (rhodopsin).

Data Fig. 1d), which suggests that β arr1 interacts with the 7TM core of M2Rpp only in a lipid bilayer¹⁵.

We optimized M2Rpp- β arr1 complexes for cryo-EM using radioligand and competition binding experiments and negative-stain electron microscopy¹⁶. The final complex—iperoxo/LY2119620-liganded M2Rpp in MSP1DIE3 lipid nanodiscs, bound to β arr1 and the antibody fragment Fab30—was monodisperse and displayed an affinity for the agonist that was 350-fold greater than that of the receptor alone (Extended Data Fig. 1e, f). Fab30 was included to stabilize the β arr1-V2Rpp interaction and to aid in cryo-EM particle alignment⁷. We chose MSP1DIE3 nanodiscs with a diameter of 12 nm, because preliminary cryo-EM analysis of an M2Rpp- β arr1 complex in smaller MSP1DH5 nanodiscs (around 9-nm diameter) indicated a mixture of ‘hanging’ conformations, in which β arr1 interacted only with V2Rpp, and ‘core’ conformations, in which β arr1 also engaged the 7TM bundle¹⁷ (Extended Data Fig. 1g, h). We previously observed such conformations in our electron microscopy analysis of complexes formed by β arr1 and the β_2 -adrenergic receptor¹⁸. Low-resolution cryo-EM maps of MSP1DH5 M2Rpp- β arr1 indicated conformational variability even among particles adopting the core interaction, with β arr1 ‘rocking’ with respect to M2Rpp and the β arr1 C-edge occasionally approaching the lipid bilayer (Extended Data Fig. 1h). Given that hydrophobic residues in the C domain of visual arrestin interact with the membrane¹⁹, we proposed that the extended lipid surface of larger MSP1DIE3 nanodiscs could stabilize complexes for studies at higher resolution.

In complexes in 12-nm nanodiscs, the interactions of β arr1 with the receptor 7TM core and the lipid surface were more uniform (Extended Data Fig. 1i). Nevertheless, variability in the size and tilt of the lipid nanodisc relative to the β arr1-Fab30 density presented a challenge for structure determination. Low-resolution analysis confirmed small differences among particle classes in the angle of β arr1 relative to the receptor 7TM, indicating inherent flexibility (Extended Data Fig. 2a). Through three-dimensional (3D) classification we identified a class of approximately 145,000 particle projections (17.4% of 831,443 particles

with a well-defined β arr1-Fab30 region) that showed solid density for the 7TM portion in one position. Other 3D classes lacked defined receptor density, possibly due to a combination of specimen distortion at the air/water interface during cryo-sample preparation²⁰ and potential projection misalignments due to the large nanodiscs. 3D refinement of the well-defined class enabled us to obtain a 3D map of the M2Rpp- β arr1 complex with a global indicated resolution of 4 Å (Fig. 1d, Extended Data Fig. 2b). This map showed relatively lower resolution in the 7TM region, including poor density of the highly mobile transmembrane helix 1 (TM1), but nevertheless enabled a confident comparison with M2R active-state structures. Subsequent refinement focused on β arr1 and its interface with M2Rpp, and yielded a map of 3.6 Å resolution displaying well-resolved features, which were used for model building in this region (Extended Data Fig. 3).

Topography of the M2R- β arr1 complex

The cryo-EM structure of the M2Rpp- β arr1 complex in nanodiscs reveals a multimodal interaction network, in which the N domain of β arr1 engages the phosphorylated C terminus of the receptor, the inter-domain loops of β arr1 engage the 7TM core and ICL2 of the receptor, and the C domain of β arr1 engages the phospholipid bilayer (Fig. 2a–c). Topologically, the central crest region interacting with the 7TM bundle occupies a similar position to the Ras domain of G proteins. Comparison of the structure of M2Rpp- β arr1 with our recent structure of the M2R- G_{α} complex (PDB: 6OIK) reveals that the 7TM bundle of the receptor adopts similar active-state configurations in both complexes, characterized by the opening of the TM6 cytoplasmic portion³ (Fig. 2d). Relatively small shifts in the cytoplasmic ends of TM5, TM6 and TM7 are probably required to accommodate β arr1, although we cannot exclude effects arising from the use of lipidic (M2Rpp- β arr1) compared with detergent (M2R- G_{α}) environments. The similarity of the transducer-bound conformations of M2R in the two structures is consistent with the similar allosteric enhancement of agonist binding to M2R by β arr1

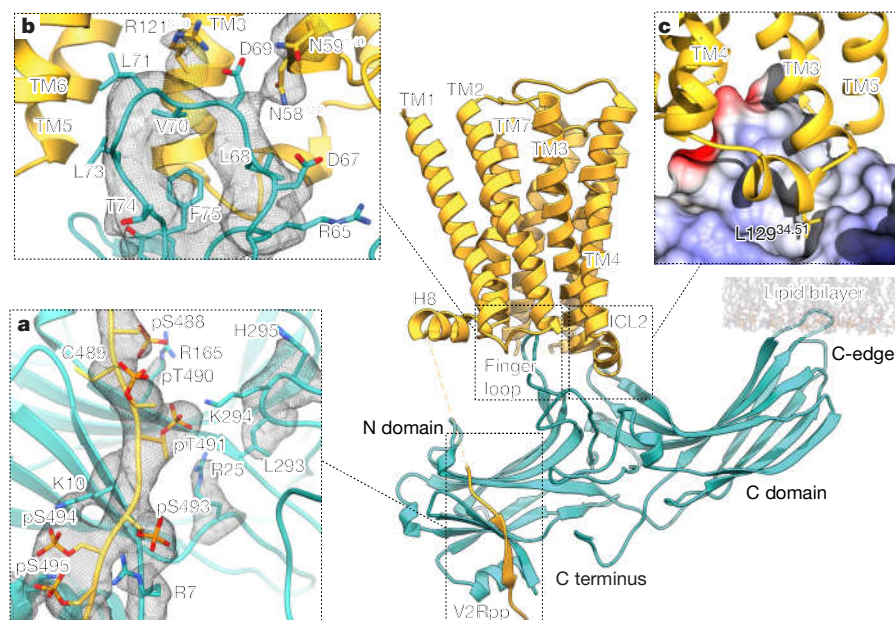


Fig. 3 | Regions of interaction between M₂R and β arr1. The M2Rpp– β arr1 complex, with dashed boxes indicating the main interaction sites between β arr1 (teal) and receptor (orange). **a**, Enlarged view of the finger loop of β arr1 inserting into the TM bundle of M2R. The model includes side-chain positions from PHENIX refinement with OPLS3e electrostatics. **b**, Enlarged view of the N domain of β arr1 bound to the C terminus of phosphorylated M2Rpp.

c, Expanded view of the interaction between ICL2 of M2R (ribbon) and a hydrophobic cleft in β arr1 (rendered as electrostatic surface; red, blue and white graduations indicate negative, positive and neutral surface potential, respectively). The mesh in all panels depicts the 3.6 Å cryo-EM map contoured at $\sigma = 3.0$ (where σ is the root mean square of the electron density in the unit cell) with a masked 2.0 Å zone around the atoms depicted.

and G protein¹⁴. In M2R–G_o, opening of the cytoplasmic portion of TM6 accommodates the α 5-helix of G_o. Similarly, displacement of TM6 accommodates the finger loop of β arr1, which occupies a similar position to the α 5-helix of the Ras domain of G_o, albeit at a slightly lower depth and involving a smaller interaction surface (Fig. 2d). This difference probably accounts for the lower affinity and inherent conformational variability of the β arr1–7TM interaction compared with the M2R–G_o interaction.

In the cryo-EM structure of M2Rpp– β arr1 and the crystal structure of rhodopsin–visual arrestin (denoted Rho–Arr1; PDB: 5WOP⁸, the arrestins show notable differences despite their similar orientations relative to the 7TM bundles (Fig. 2e). β arr1 is tilted approximately 7° further towards the membrane (Extended Data Fig. 4a)—this is probably due to the interaction of the C domain with the nanodisc phospholipids, although crystal packing interactions in the structure of Rho–Arr1 might also contribute to these differences. In addition, the finger loops display differences in both positioning and structure. Whereas the finger loop of visual arrestin is modelled with an α -helical segment (residues E71–G77), the finger loop of β arr1 adopts an extended loop configuration. The other arrestin interdomain loops adopt similar conformations in these two GPCR–arrestin structures (Extended Data Fig. 4b).

Interaction of β arr1 with M2R

β arr1 interacts with M2Rpp through interfaces with the phosphorylated C terminus, the 7TM core, and ICL2 of the receptor (Fig. 3). The phosphopeptide– β arr1 interaction is essentially identical to that in our crystal structure of V2Rpp– β arr1–Fab30⁷, and six phosphorylated residues are well resolved in the 3.6 Å map. The peptide binds to a positively charged crevice on the N domain by displacing the C terminus of β arr1, which destabilizes the arrestin polar core and enables the gate loop to flip towards the N domain. A critical phosphorylated residue on the C terminus of the receptor, T491 (T360 in V2Rpp), establishes interactions with R25 (N domain) and K294 (gate loop) of β arr1,

stabilizing an activated conformation characterized by interdomain twisting²¹ (Fig. 3a). The fused V2Rpp is 36 residues long and provides ample flexibility so as not to limit the orientation of arrestin; this is also evident by the lack of observed density for the 23 residues that do not engage arrestin. Similarly, although we cannot rule out the possibility that natively phosphorylated ICL3 residues influence the relative orientation of β arr1, this seems unlikely given the length of this loop (152 residues) and its lack of order in our structure.

The second interface between the finger loop of β arr1 and the 7TM bundle of the receptor involves both hydrophobic and electrostatic interactions (Fig. 3b). The C-terminal finger loop region—including residues L68, V70, L71 and F75—is hydrophobically packed against the side of the receptor pocket formed by TM3, TM5, TM6 and ICL2. On the opposite side, the map of the N-terminal part of the finger loop displays partial side-chain densities of ionic residues R65, E66 and D67, revealing both their positions and their relative mobilities. A potential salt bridge between R65 (finger loop) and D135 (middle loop) of β arr1 could stabilize the conformation of the finger loop, whereas E66—although resolved only through C β —points towards R62 of the finger loop and K138 of the middle loop (Extended Data Fig. 5a). At the tip of the finger loop, D69 seems to be positioned to form hydrogen bonds or salt bridges with several receptor residues, including N58^{2,39} and R121^{3,50} of the highly conserved DRY motif (superscripts denote Ballesteros–Weinstein numbering for GPCRs²²). These interactions formed consistently across multiple, independent molecular dynamics simulations, suggesting that this interaction network stabilizes the M2Rpp– β arr1 complex (Extended Data Fig. 5b, c). Considering that R^{3,50}–E^{6,30} salt bridges stabilize the inactive receptor state of M2R and other Class A GPCRs, β arr1 may stabilize an active receptor partly by engaging R^{3,50}. The mutant β arr1(D69A) displays significantly reduced coupling to M2Rpp, which supports this hypothesis (Extended Data Fig. 5d).

The finger-loop fold, the interdomain twist, and the binding mode through the C terminus of V2Rpp that we observe here for β arr1 are very similar to those in the crystal structure of the active β arr1–V2Rpp–Fab30⁷ (PDB: 4JQI, Extended Data Fig. 4c). However, the crystal

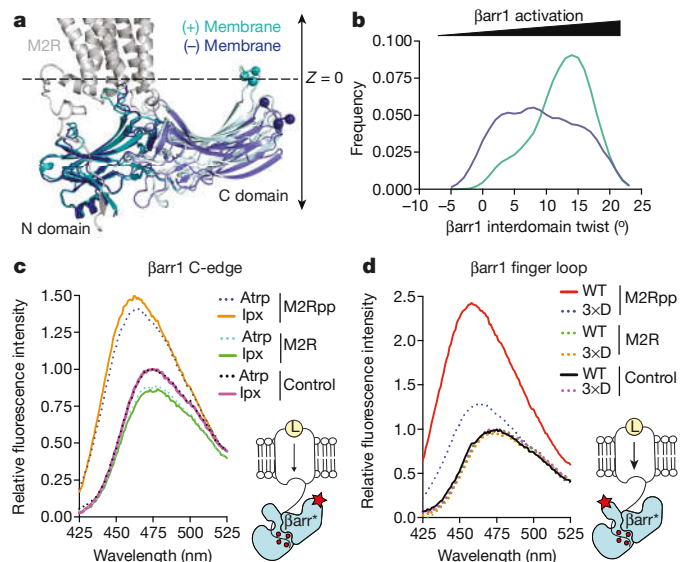


Fig. 4 | Anchoring of the C domain of β arr1 to the lipid membrane.

a, b. Atomistic simulations of β arr1 in complex with M2Rpp (in the absence of Fab30) to investigate the position of the C-edge of β arr1 (**a**) and the β arr1 interdomain twist angle in the absence (dark blue) or presence (cyan) of a lipid membrane (dashed black line in **a**) (**b**). **c.** HDL-M2Rpp, but not HDL-M2R, enhances the fluorescence of bimane at the C-edge of β arr1 (inset, red star) independent of the presence of antagonist (atropine) or agonist (iperoxide). **d.** The ability of iperoxide-activated HDL-M2Rpp to increase the fluorescence of bimane in the finger loop of β arr1 (inset, red star) requires receptor phosphorylation (compare with HDL-M2R) and is markedly reduced by the C-edge 3xD mutations (L335D/L338D/S340D). Data are representative of three independent experiments.

structure shows a large twist in the orientation of the finger loop, due to the N-terminal segment of the phosphopeptide inducing a twist in the flanking β -strands S5 and S6. This V2Rpp segment must disengage before the finger loop can be inserted into M2Rpp and is not ordered in our map. The positioning of the N-terminal V2Rpp segment in the crystal structure could be due to lattice packing. Alternatively, these differences could reflect a multi-step binding mechanism, in which the initial binding of β arr1 to the full length of the phosphopeptide is followed by release of the N terminus of the phosphopeptide, flipping of the β arr1 finger loop, and burial of the hydrophobic residues of the finger loop in the intracellular cavity of the receptor (Extended Data Fig. 6a).

The M2Rpp ICL2 rests in a cleft between the N- and C domains of β arr1, which comprises portions of the finger, middle, gate and C-loops. Although this portion of the map lacks well-resolved side-chain features, the density suggests that ICL2 adopts the same helical conformation as in M2R-G_o (Fig. 3c). This ICL2 orientation places L129^{34,51} facing a β arr1 hydrophobic cleft that is capable of accommodating even larger side chains (Extended Data Fig. 6b). Although leucine is most prevalent at this position in other class A GPCRs that couple to β -arrestins, larger hydrophobic amino acids (such as methionine and phenylalanine) are common (Extended Data Fig. 6c), suggesting one source of the promiscuity of β -arrestins. Notably, the same hydrophobic ICL2 residue often engages in van der Waals interactions with G proteins^{3,23–25}.

Membrane interaction of β arr1 C domain

The cryo-EM maps reveal that the C-edge of β arr1 interacts with the nanodisc, with one C-edge loop contacting the membrane (loop 1, residues 191–196, LMSDKP) and a second loop burying itself in the lipid bilayer (loop 2, residues 330–340, SRGGLGLAS) (Extended Data Fig. 7). Interaction of the C-edge of arrestin with the membrane has

been suggested by the structure of Rho-Arr1, and by site-specifically labelling the C-edge of visual arrestin with environmentally sensitive probes^{12,19}. Given that the affinity of β arr1 for the 7TM core of M2Rpp is higher in nanodiscs than in detergent (Fig. 1b, c), we proposed that there are functional consequences to the C-edge–lipid interaction. First, we performed multiple, independent molecular dynamics simulations²¹ with M2Rpp– β arr1 initially embedded in either a continuous membrane (enabling C-edge interaction) or a 9-nm-diameter nanodisc (too small for C-edge interaction) (Fig. 4a, Extended Data Fig. 8a). In four of the five simulations using the full membrane, β arr1 remained in its active conformation with an average interdomain twist angle of around 17°—similar to the interdomain twist observed in the crystal structure of the β arr1-V2Rpp-Fab30 complex⁷. However, in small nanodiscs, β arr1 adopted a more heterogeneous conformational ensemble dominated by inactive conformations (with an interdomain twist angle of between 0° and 7°) (Fig. 4b); this suggests that the interaction between the C-edge and the membrane stabilizes the active conformation of β arr1.

We site-specifically labelled C-edge loop 2 with mBr (L338mBr) to assess whether the β arr1–lipid interaction requires the binding of β arr1 to the M2Rpp phosphorylated tail or the 7TM bundle¹⁹. An increase in fluorescence of β arr1 L338mBr, which reflects insertion of the C-edge into a hydrophobic environment, requires M2Rpp phosphorylation and nanodisc reconstitution (Fig. 4c, Extended Data Fig. 8b). This is consistent with previous observations that the engagement of visual arrestin with membranes requires rhodopsin phosphorylation¹⁹. However, the change in β arr1 L338mBr fluorescence is not agonist (iperoxide)-dependent, which indicates that the C-edge–lipid interaction does not require interaction between β arr1 and the 7TM bundle. Because the C-edge–membrane interaction is observed in nanodiscs but not in detergent (Fig. 1b, c), we suggested that it might facilitate β arr1–7TM coupling. We sought to test this by attenuating the hydrophobicity of the C-edge with aspartic acid substitutions at L335, L338 and S340 of β arr1 (β arr1(3xD))¹⁹. Fluorescence of β arr1 in which the finger loop was labelled with bimane (Fig. 4d, Extended Data Fig. 8c) confirmed that the 3xD mutation markedly reduced the interaction of β arr1 with the 7TM bundle of HDL-M2Rpp. Thus, maximal coupling of β arr1 to the 7TM bundle of M2Rpp requires not only receptor phosphorylation but also interaction of the C-edge of β arr1 with the phospholipid bilayer.

Furthermore, we investigated whether the interaction between the C-edge of β arr1 and the membrane modulates β arr1-mediated receptor desensitization and internalization. We measured the ability of purified wild-type β arr1 and β arr1(3xD) to inhibit (desensitize) HDL-M2Rpp-mediated G-protein activation in vitro. Iperoxide-induced activation of HDL-M2Rpp increases the GTPase activity of purified heterotrimeric G_i. Whereas wild-type β arr1 blocks 50% of G_i activity, β arr1(3xD) has almost no effect (Fig. 5a). We then compared the ability of GFP-tagged wild-type β arr1 (GFP– β arr1(WT)) and GFP– β arr1(3xD) to induce the internalization of Flag-tagged M2R (Flag-M2R) in β arr1/ β arr2-null HEK293 cells. After 5 minutes of iperoxide stimulation, GFP– β arr1(WT) translocates from the cytoplasm to the plasma membrane. After 30 minutes, both GFP– β arr1(WT) and Flag-M2R (more than 40%) are largely internalized (Fig. 5b, c, Extended Data Fig. 9a, b). By contrast, GFP– β arr1(3xD)—which is expressed at similar levels to the wild-type (Extended Data Fig. 9c, d)—remains in the cytoplasm even after 30 minutes, with little Flag-M2R internalization (10%). Notably, the impaired recruitment of GFP– β arr1(3xD) to wild-type M2R and V2R (Extended data Fig. 9e) shows that membrane anchoring contributes to the recruitment of β arr1 to natively phosphorylated GPCRs, not only the engineered M2Rpp construct. Thus, the interaction of β arr1 with the membrane is critical for desensitization and internalization of the receptor.

Three-site interaction model

The binding of arrestin to GPCRs has been viewed as a two-part process: binding to the phosphorylated ICLs or the C terminus of the receptor

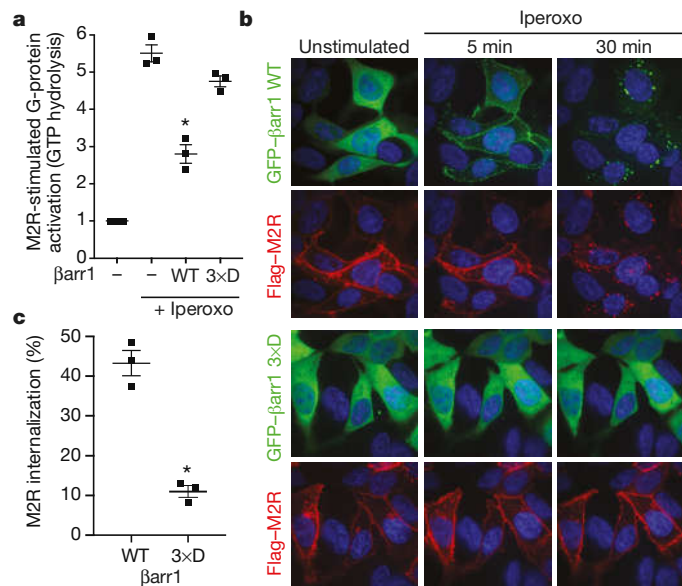


Fig. 5 | The functionality of β arr1 depends on C domain–lipid interactions.

a, Activation of purified heterotrimeric G_i protein by iperoxo-stimulated HDL-M2Rpp *in vitro* is reduced by wild-type β arr1 but not by β arr1(3xD). Data are the mean of three independent experiments with error bars representing s.e.m. * $P < 0.0001$, one-way ANOVA compared to wild-type β arr1 plus iperoxo.

b, Iperoxo stimulation of Flag-M2R causes plasma membrane recruitment of GFP- β arr1(WT) and subsequent receptor internalization, which is impaired by the 3xD mutations, as demonstrated by confocal microscopy. β arr1, nuclei and M2R are coloured green, blue, and red, respectively. Confocal microscopy images are representative of three independent experiments. **c**, Quantification of Flag-M2R internalization by flow cytometry in same cells as **b**. Data are the mean of three independent experiments with error bars representing s.e.m. * $P = 0.0008$, unpaired two-sided t -test compared to wild-type β arr1.

induces conformational changes in β -arrestin, which subsequently promotes its coupling to the 7TM bundle of the receptor. Our findings, supported by observations of the Rho-Arr1 complex^{12,19} and the NTSR1- β arr1 complex²⁶, expand this model to include a critical β arr1–lipid interaction (Extended Data Fig. 9f). In this three-site interaction model, the recruitment of arrestin requires both GPCR phosphorylation and an interaction of β arr1 with the plasma membrane (Fig. 5b), which could increase the arrestin concentration at the cell membrane before receptor activation. The C-edge–membrane interaction subsequently enhances the binding of arrestin to the 7TM bundle (Fig. 4d) and the desensitization of G-protein activation (Fig. 5a).

Whereas membrane anchoring may be a conserved property of arrestins, the mechanisms that underlie membrane association could be distinct. The C-edge loop buried in the nanodisc is found in visual arrestin and the dominant form of β arr1, but not in β arr2 or an alternative β arr1 splice variant. The binding site of phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂)—a phospholipid observed in the structure of the NTSR1- β arr1 complex, described in a related study²⁶—is conserved in β -arrestins but not in retinal arrestins^{12,19,27,28} (Extended Data Fig. 7). Accordingly, the variability of membrane compositions could provide yet another level of regulation for GPCR desensitization, internalization and signalling. Notably, β arr2 can stimulate MAP kinase signalling from clathrin-coated structures after dissociation from activated GPCRs²⁹. These and other recent data suggest that the lipid membrane stabilizes the activation of β -arrestin when bound to, and even after dissociation from, the receptor^{21,29–31}. Thus, as is evident from the structures of β arr1 bound to M2Rpp or NTSR1²⁶, a complex cooperative network of low-affinity interactions involving both receptor and membrane phospholipids endow arrestins with the necessary plasticity to variably couple to hundreds of GPCRs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1954-0>.

- Rajagopal, S. & Shenoy, S. K. GPCR desensitization: Acute and prolonged phases. *Cell Signal.* **41**, 9–16 (2018).
- Reiter, E., Ahn, S., Shukla, A. K. & Lefkowitz, R. J. Molecular mechanism of β -arrestin-biased agonism at seven-transmembrane receptors. *Annu. Rev. Pharmacol. Toxicol.* **52**, 179–197 (2012).
- Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G. & Koblika, B. K. Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science* **364**, 552–557 (2019).
- Gurevich, V. V. & Gurevich, E. V. GPCR signaling regulation: the role of GRKs and arrestins. *Front. Pharmacol.* **10**, 125 (2019).
- Chen, Q., Iverson, T. M. & Gurevich, V. V. Structural basis of arrestin-dependent signal transduction. *Trends Biochem. Sci.* **43**, 412–423 (2018).
- Scheerer, P. & Sommer, M. E. Structural mechanism of arrestin activation. *Curr. Opin. Struct. Biol.* **45**, 160–169 (2017).
- Shukla, A. K. et al. Structure of active β -arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature* **497**, 137–141 (2013).
- Zhou, X. E. et al. Identification of phosphorylation codes for arrestin recruitment by G protein-coupled receptors. *Cell* **170**, 457–469.e413 (2017).
- Miller, W. E. & Lefkowitz, R. J. Expanding roles for beta-arrestins as scaffolds and adaptors in GPCR signaling and trafficking. *Curr. Opin. Cell Biol.* **13**, 139–145 (2001).
- Smith, J. S., Lefkowitz, R. J. & Rajagopal, S. Biased signalling: from simple switches to allosteric microprocessors. *Nat. Rev. Drug Discov.* **17**, 243–260 (2018).
- García-Nafria, J. & Tate, C. G. Cryo-EM structures of GPCRs coupled to G_{12} , G_i and G_{13} . *Mol. Cell. Endocrinol.* **488**, 1–13 (2019).
- Kang, Y. et al. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–567 (2015).
- Kruse, A. C. et al. Muscarinic acetylcholine receptors: novel opportunities for drug development. *Nat. Rev. Drug Discov.* **13**, 549–560 (2014).
- Staus, D. P. et al. Sortase ligation enables homogeneous GPCR phosphorylation to reveal diversity in β -arrestin coupling. *Proc. Natl Acad. Sci. USA* **115**, 3834–3839 (2018).
- Gurevich, V. V., Pals-Rylaarsdam, R., Benovic, J. L., Hosey, M. M. & Onorato, J. J. Agonist-receptor-arrestin, an alternative ternary complex with high agonist affinity. *J. Biol. Chem.* **272**, 28849–28852 (1997).
- Peisley, A. & Skiniotis, G. 2D projection analysis of GPCR complexes by negative stain electron microscopy. *Methods Mol. Biol.* **1335**, 29–38 (2015).
- Grinkova, Y. V., Denisov, I. G. & Sligar, S. G. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng. Des. Sel.* **23**, 843–848 (2010).
- Shukla, A. K. et al. Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature* **512**, 218–222 (2014).
- Lally, C. C., Bauer, B., Selent, J. & Sommer, M. E. C-edge loops of arrestin function as a membrane anchor. *Nat. Commun.* **8**, 14258 (2017).
- Noble, A. J. et al. Routine single particle cryoEM sample and grid characterization by tomography. *eLife* **7**, e34257 (2018).
- Latorraca, N. R. et al. Molecular mechanism of GPCR-mediated arrestin activation. *Nature* **557**, 452–456 (2018).
- Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* Vol. **25** (ed. Sealfon, S. C.) 366–428 (Academic, 1995).
- Koehl, A. et al. Structure of the μ -opioid receptor– G_i protein complex. *Nature* **558**, 547–552 (2018).
- Rasmussen, S. G. et al. Crystal structure of the β_2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
- Krishna Kumar, K. et al. Structure of a signaling cannabinoid receptor 1–G protein complex. *Cell* **176**, 448–458.e12 (2019).
- Huang, W. et al. Structure of the neurotensin receptor 1 in complex with β -arrestin 1. *Nature* (in press).
- Gaidarov, I., Krupnick, J. G., Falck, J. R., Benovic, J. L. & Keen, J. H. Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. *EMBO J.* **18**, 871–881 (1999).
- Parruti, G. et al. Molecular analysis of human beta-arrestin-1: cloning, tissue distribution, and regulation of expression. Identification of two isoforms generated by alternative splicing. *J. Biol. Chem.* **268**, 9753–9761 (1993).
- Eichel, K., Jullié, D. & von Zastrow, M. β -Arrestin drives MAP kinase signalling from clathrin-coated structures after GPCR dissociation. *Nat. Cell Biol.* **18**, 303–310 (2016).
- Eichel, K. et al. Catalytic activation of β -arrestin by GPCRs. *Nature* **557**, 381–386 (2018).
- Nuber, S. et al. β -Arrestin biosensors reveal a rapid, receptor-dependent activation/deactivation cycle. *Nature* **531**, 661–664 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Receptor expression and purification

M2R containing an N-terminal Flag tag, C-terminal sortase ligation consensus sequence (LPETGGH), and 6×His-tag was cloned into pcDNA-zeo-teto and stably expressed in tetracycline-inducible Expi293F cells (Invitrogen)¹⁴. Cells were not authenticated or routinely tested for mycoplasma. Cells were grown to a density of 4–5 million cells per ml and then treated with Expi293F expression enhancers (Thermo Fisher), 5 μ M atropine, and 5 μ M kifunensine to restrict glycosylation. Expression was induced 18 h thereafter by addition of 4 μ g ml⁻¹ doxycycline and 5 mM sodium butyrate. Cells were collected 48 h post-induction and stored at -80 °C until further processing. Flag-M2R-LPETGG-His6 was purified as previously described^{14,32}, with all purification steps conducted at 4 °C with protease inhibitors (benzamidine and leupeptin) unless stated otherwise. In brief, cells were lysed for 30 min by stirring in 10 ml g⁻¹ wet cell mass of lysis buffer (10 mM Tris (pH 7.4), 2 mM EDTA, and 10 mM MgCl₂, 5 units per ml benzonase, 5 μ M atropine, and 2 mg ml⁻¹ iodoacetamide). Membrane was pelleted at 30,000g for 20 min and resuspended in 10 ml g⁻¹ original cell pellet mass of solubilization buffer (20 mM HEPES (pH 7.4), 750 mM NaCl, 1% *n*-dodecyl- β -D-maltoside (DDM), 0.05% cholesterol hemisuccinate, 10% glycerol, 5 units per ml benzonase, 5 μ M atropine, and 2 mg ml⁻¹ iodoacetamide). After extensive homogenization using a Dounce homogenizer, solubilizing membrane was sequentially stirred at room temperature and 4 °C for 1 h each. Insoluble material was removed by centrifugation at 30,000g for 30 min, and the supernatant was loaded onto M1-Flag resin with 2 mM CaCl₂ at 1–3 ml min⁻¹. M1-Flag resin was washed with 5 column volumes each of wash buffer (20 mM HEPES (pH 7.4), 0.1% DDM, 0.01% CHS, 2 mM CaCl₂, and 1 μ M atropine) containing high (750 mM) and low (100 mM) NaCl at ratios of 4:0, 3:1, 2:2, 1:3 and 0:4, respectively. Receptor was eluted in elution buffer (20 mM HEPES (pH 7.4), 100 mM NaCl, 0.1% DDM, 0.01% CHS, 0.2 mg ml⁻¹ Flag-peptide, 1 μ M atropine, and 5 mM EDTA), and glycosylation was removed by incubation with a 1:10 protein ratio of EndoH to M2R for 90 min at room temperature. The enzyme sortase was used to ligate the synthetic phosphopeptide GGG-V2Rpp (GGG-ARGRpTPPPSLGPQDEpSCpTpTapSpSpSLAKDTSS) on the C terminus of M2R (M2Rpp) as previously described¹⁴. Monomeric receptor was collected by size-exclusion chromatography on a Superdex200 Increase column (GE Healthcare Life Sciences).

HDL reconstitution

DDM-M2R was pre-incubated with twofold molar excess atropine on ice for 30 min before reconstitution. Receptor (5 μ M) was mixed with 8 mM POPC/POPG lipids and 150 μ M MSP1D1E3 on ice for 1 h in a final buffer composition of 20 mM HEPES (pH 7.4), 100 mM NaCl, and 0.5 mM EDTA. Detergent was removed using Bio-Beads (BioRad) (50 mg per 100 μ l reaction volume) and rotated overnight at 4 °C. The supernatant was separated from the Bio-Beads using a 28-gauge needle and was diluted with 20 mM HEPES (pH 7.5) and 100 mM NaCl (HN buffer) to obtain a final concentration of 2 μ M HDL-M2R. HDL-M2R was rotated with M1-Flag resin for 30 min at room temperature and then an additional 15 min after adding 2 mM CaCl₂. M1-Flag resin was quickly washed in column format with five resin volumes of HN buffer with 2 mM CaCl₂. HDL-M2R was eluted with HN buffer containing 0.2 mg ml⁻¹ Flag peptide and 5 mM EDTA. Size-exclusion chromatography was used to collect monomeric HDL-M2R and to remove Flag peptide.

β arr1 and Fab30 purification

To enhance expression and stability, a minimal cysteine (C59A, C125S, C140I, C150V, C242V, C251V and C269S) and truncated (after amino

acid 393) variant of rat β -arrestin1 (β arr1-MC-393) in pGEX4T was generated. β arr1-MC-393 was expressed and purified as previously described³³. In brief, GST- β arr1-MC-393 was expressed in BL21(DE3) bacteria, lysed by sonication, and captured using glutathione (GST) Sepharose. β arr1-MC-393 was cleaved from GST by thrombin digestion and further purified using HiTrap Q Sepharose anion exchange. Fab30 was purified as described previously⁷.

HDL-M2Rpp β arr1 complex formation and purification

HDL-M2Rpp was pre-incubated for 30 min on ice with a fivefold molar excess of iperoxo and LY211960 and then 90 min at room temperature with a twofold molar excess of β arr1-MC-393 and Fab30. The complex of HDL-M2Rpp- β arr1-Fab30 was separated from unbound β arr1-MC-393 and Fab30 by M1-Flag chromatography as described above. After M1-Flag elution, the complex was subjected to size-exclusion chromatography on a Superdex200 Increase column with HN containing 1 μ M iperoxo and 2 μ M LY211960. Peak fractions were concentrated to obtain a final concentration of 2 mg ml⁻¹ for cryo-EM analysis.

Co-immunoprecipitation

DDM-M2R or DDM-M2Rpp (4 μ g) was pre-incubated with iperoxo (10 μ M) for 30 min on ice before the addition of twofold molar excess purified β arr1-MC-393 and Fab30. Reactions were incubated at room temperature for 1 h, rotated with M1-Flag resin for 20 min, washed with HN buffer containing 0.1% DDM, and eluted with HN buffer with 0.1% DDM, 0.4 mg ml⁻¹ Flag peptide and 10 mM EDTA. Proteins were separated by SDS-PAGE and visualized by Instant Blue stain (Expedeon)

Radioligand binding assays

Competition equilibrium binding experiments with DDM-M2R were conducted using a scintillation proximity assay (Perkin Elmer) in a final volume of 100 μ l containing HN buffer with 0.1% DDM and 1 mg ml⁻¹ BSA. DDM-M2Rpp was incubated for 1 h at room temperature with 1 μ M β arr1 or 10 μ M LY211960, 1 nM [³H]N-methylscopolamine (NMS), and varying concentrations of iperoxo. Yttrium silicate (YSi) protein A beads (Perkin Elmer) coated with Flag-M1 antibody were subsequently added for 30 min, and YSi emission was read on a Wallac 1450 Microbeta Plus. Radioligand binding with HDL-M2R was conducted in HN with 1 mg ml⁻¹ BSA with the same final concentrations of components as above. Reactions proceeded for 90 min and were collected onto glass fibre filters (GF-B) soaked with 0.3% polyethyleneimine using a Brandel harvester. Data from competition radioligand binding assays were analysed using GraphPad software.

Bimane fluorescence

Purified rat β arr1-MC-393(V70C), β arr1-MC-393(L337C) and β arr1-MC-393(V70C/L334D/L337D/S339D) (3×D) were labelled overnight at 4 °C with a threefold molar excess of mBr (Sigma) and an additional threefold molar excess added for 1 h at room temperature the next day. Reactions were quenched with L-cysteine, and free mBr was removed by size-exclusion chromatography. For bimane experiments, DDM-M2R or HDL-M2R was pre-incubated with 10 μ M iperoxo or atropine and twofold molar excess Fab30 for 20 min at room temperature in HN buffer containing 1 mg ml⁻¹ BSA. Bimane-labelled complexes mBr- β arr1-MC-393(V70C), mBr- β arr1-MC-393(L337C) or mBr- β arr1-MC-393(V70C/3×D) were added to a final concentration of 200 nM with a 1.5-fold molar excess of DDM- or HDL-M2R. For DDM-M2R, the buffer additionally contained 0.1% DDM and 0.01% CHS. Reactions were equilibrated for 30 min in black, solid-bottom 96-well microplates (Corning) before fluorescence emission spectra were collected on a CLARIOstar plate reader (BMG Labtech) in top-read mode, with excitation at 370 nm (16-nm bandpass) and emission scanning from 400 nm to 600 nm (10-nm bandpass) in 1-nm increments. Wells for background subtraction contained all components except bimane-labelled β arr1-MC-393.

Article

Statistics were conducted by comparing the area under the curves (GraphPad Prism) from at least three independent experiments.

Cryo-EM data acquisition and data processing

Before cryo-EM preparation, all samples were screened for quality using conventional negative-stain electron microscopy¹⁶. In preliminary cryo-EM experiments, M2Rpp- β arr1-Nb24-scFv30 (Nb24: camelid nanobody) complex in the smaller (~9-nm diameter) MSP1D1H5 nanodiscs, at a concentration of 1 mg ml⁻¹, was applied to glow-discharged 200-mesh grids (Quantifoil R1.2/1.3), and vitrified using a Vitrobot Mark IV (Thermo Fisher Scientific) at 4 °C and 100% humidity. Images were collected on a Titan Krios equipped with a Gatan K2 Summit direct electron camera, at a nominal magnification of 59,000 \times with a pixel size of 0.86 Å. A total of 3,589 movie stacks were acquired with dose-fractionation into 40 frames with an accumulated dose of 70 electrons per Å⁻². Dose-fractionated stacks were subjected to beam-induced motion correction using MotionCor2³⁴. Data processing was performed in RELION3.0³⁵. Particle projections (487,582) were selected by template-based autopicking and subjected to 2D classification. A subset of around 70,000 particles showing density for both HDL-receptor and β arr1 were subjected to 3D classification. One out of six 3D classes showed a 'hanging' arrestin conformation, and three classes showed a 'core' interaction between receptor and arrestin but with 'rocking' motion of arrestin relative to the receptor-embedded nanodisc (Extended Data Fig. 1e, f).

M2Rpp- β arr1-Fab30 complex in the larger (~12-nm diameter) MSP1D1E3 nanodiscs was prepared as described above, except that cryo-EM images were collected on a Titan Krios equipped with the Gatan GIF Quantum LS Imaging energy filter using a Gatan K2 Summit direct electron camera in counted mode, corresponding to a pixel size of 1.06 Å. Each image was dose-fractionated into 40 frames with a dose rate of 7 electrons per pixel per second and total exposure time of 8 s, resulting in an accumulated dose of 50 electrons per Å⁻². Data were collected in seven independent sessions resulting in a total number of 30,454 movie stacks. Contrast transfer function parameters for each micrograph were determined by Gctf v1.06³⁶. Auto-picking was performed with templates that were generated from manually picked particles. Particles were subsequently screened by 2D reference-free classification followed by 3D classification. Particle projections were pooled together after independent 3D classification of each dataset. A representative data-processing workflow is shown in Extended Data Fig. 2b. In 2D and 3D classification, around 25% of particle projections showing well-defined density for complex components indicated an extra density connected to Fab30, which we suspect to be a non-specifically attached second Fab30. Because this domain was variable among complexes, a mask was applied to remove its density during 3D classification. Particles from each dataset contributing to 3D reconstructions with well-defined features were merged with the other six subsets obtained in a similar strategy, resulting a total number of 831,443 projections with well-defined particle components. Owing to the variability in size and tilt of the large MSP1D1E3 nanodisc, which was dominating the projection alignment, the 831,443 particle projections were subjected to a focused alignment on β arr1-Fab30 density followed by 3D classification without alignment while masking out the nanodisc density. One 3D class accounting for 145,618 particles showed a well-defined transmembrane portion, and its particle projections were subjected to 3D refinement after subtracting from raw images the density of the nanodisc and the constant (CL/CH1) portion of Fab30 due to its relative flexibility, but leaving the scFv portion intact. This strategy resulted in a 3D reconstruction of the M2Rpp- β arr1-Fab30(scFv) with global indicated resolution of 4.0 Å. In this map, the extracellular portion of M2Rpp showed the lowest relative resolution. To improve the densities for the β arr1 and M2Rpp interface we used the refinement parameters from the global map to subtract more than half of the extracellular receptor side together with nanodisc and the

constant (CL/CH1) domains of Fab30 from raw particle images. The subtracted particles were imported to cisTEM³⁷ for a refinement with local search, resulting in a focused map with global indicated resolution of 3.6 Å. Global resolution estimations were obtained with PHENIX from two half maps at Fourier shell correlation (FSC) cutoff of 0.143. Local resolution determination of the global M2Rpp- β arr1-Fab30(scFv) map and the focused interface M2Rpp- β arr1-Fab30(scFv) map was calculated with blocres from the Bsoft package³⁸ at an FSC cutoff of 0.5 (Extended Data Figs. 3b, c, Extended Data Table 1).

Model building and refinement

The initial model was prepared by manually docking the cryo-EM structure of active-state M2R (PDB: 6OIK) and the crystal structure of activated V2Rpp- β arr1-Fab30 (PDB: 4JQI) into the cryo-EM map. We used the global M2Rpp- β arr1-Fab30(scFv) 4 Å map for modelling the 7TM bundle and the focused interface M2Rpp- β arr1-Fab30(scFv) 3.6 Å map for modelling β arr1 and its interface with M2R. Iterative rounds of real-space refinement were performed with *phenix.real_space_refine* in PHENIX³⁹ and manual model building with Coot⁴⁰. Rosetta was periodically used to assess whether more optimal models existed. Additional refinement near the end of model building was performed with PHENIX using the state-of-the-art OPLS3e force field with electrostatics. A final refinement was performed on the full complex with PHENIX using the global M2Rpp- β arr1-Fab30(scFv) 4 Å map (Extended Data Table 1). Independent FSC curves for model-map correlations were calculated between the resulting model and the two maps (Extended Data Fig. 3a). Model overfitting was evaluated through its refinement against the focused interface M2Rpp- β arr1-Fab30 half maps after randomly displacing all atoms by 0.2 Å (Extended Data Fig. 3a).

Model analysis

Electrostatic potential surfaces were calculated with APBS⁴¹ in Chimera. Charges were prepared with PDB2PQR⁴² using the PARSE force field⁴³. Tilt angle of arrestin compared with receptor was calculated between residues V37 (β arr1), I317 (β arr1) and R121 (M2) in the case of M2Rpp- β arr1, and V2042 (arr1), I2324 (arr1), and R135 (rhodopsin) for the Rho-Arr1 structure.

Molecular dynamics simulations

Starting from a model of the M2R- β arr1 complex derived from an earlier refinement (model available upon request), the OPM webserver⁴⁴ was used to orient the structure with respect to the plane of the lipid bilayer. The aligned structure was then prepared further using CHARMM-GUI⁴⁵ to place the structure in either a membrane bilayer or a MSP1D1-44 nanodisc. In both cases a 3:2 ratio of POPC to POPG was used for the lipid system. The systems were hydrated with TIP3P water and charge-neutralized with 150 mM NaCl. Further system preparation was performed in VMD⁴⁶, in which palmitoylation was added to residue Cys457 of the receptor. Serines 488, 493, 494, 495 and threonines 490 and 491 were phosphorylated with dibasic phosphate. The systems were simulated in NAMD⁴⁷ with the OPLS-AA/M force field⁴⁸. OPLS-AA parameters for iperoxo and LY2119620 were obtained with the LigParGen server⁴⁹. OPLS-AA parameters for POPC were taken from ref.⁵⁰, while POPG parameters were adapted from POPC and the OPLS-AA small-molecule set. OPLS-AA/M parameters for phosphorylated serine and threonine were also developed for this work on the basis of existing OPLS-AA parameters for phosphates⁵¹. Parameters for palmitoylated cysteine were developed for this work by combining OPLS-AA/M parameters for cysteine with the lipid parameters of ref.⁵⁰. All simulations followed a 2-fs timestep in the NPT ensemble using a Langevin thermostat set at 300 K with a dampening coefficient of 1 ps⁻¹ and a Nosé-Hoover Langevin piston barostat set at 1 atm with a period of 50 fs and a decay of 25 fs. Non-bonded interactions were smoothed starting at 10 Å to 12 Å, where long-range interactions were treated with particle mesh Ewald. Systems were minimized for 2,000 steps before being slowly

heated from 0 K to 300 K in 20-K increments with 0.4 ns of simulation at each increment. During the heating phase, harmonic restraints of 1 kcal mol⁻¹ Å⁻² were applied to all lipid, protein, and small-molecule non-hydrogen atoms. The system was then equilibrated with harmonic restraints of 1 kcal mol⁻¹ Å⁻² applied to all non-hydrogen protein atoms for 2 ns, followed by 2 ns of simulation with restraints on only protein Cα atoms. These restraints were then gradually reduced first to 0.3 kcal mol⁻¹ Å⁻² for 2 ns before being removed completely. This aggregate 12 ns of heating and equilibration and an additional 18 ns of production simulation were discarded from the final calculated quantities. Each system was then simulated for five replicates of up to 200 ns to produce the results used in this work. Trajectories were analysed with the VMD software⁴⁶. Interdomain rotation was calculated as described in ref.²¹ using the inactive and active state crystal structures of arrestin (PDB: 1CF1⁵² for the inactive structure and PDB: 4ZWJ¹²) to define the axis of rotation. These structures were chosen as references, as the Rho–Arr1 structure was thought to be the closest analogue to our structure; however, results were qualitatively similar to those of other references.

Confocal microscopy

HEK293 (ATCC) cells lacking endogenous βarr1/2 as described elsewhere^{53,54} were transiently transfected with a 1:4 DNA ratio of pcDNA-teto-M₂R and GFP-βarr1(WT) or the 3×D variant, respectively. Cells were not authenticated or routinely tested for mycoplasma. Cells were split into 35-mm glass bottom microwell dishes (MatTek) and 24 h thereafter serum-starved for 2 h in MEM media containing 20 mM HEPES (pH 7.4) and 1 mg ml⁻¹ BSA. Cells were incubated with Alexa-650-labelled Flag–M1 antibody and NucBlue Live cell stain (Invitrogen) for 1 h at 37 °C, washed twice in starvation media, and imaged using confocal microscopy. In parallel, transfected cells split into 6-well plates were stimulated for 30 min at 37 °C in the presence or absence of 1 μM iperoxo. Cells were immediately placed on ice and kept at 4 °C for the remainder of the experiment. Cells were washed twice with cold phosphate-buffered saline and then detached with 0.05% EDTA. Cells were resuspended in assay buffer (Hanks balanced salt solution, 20 mM HEPES pH 7.4, 3 mM CaCl₂, 1 mg ml⁻¹ BSA) and stained with Alexa-650-labelled Flag–M1 antibody for 30 min at 4 °C. Cells were washed once with assay buffer before analysis by flow cytometry (Bio-Rad S3e Cell Sorter). Data were analysed using FlowJo software, gating for GFP-positive singlet cells.

G-protein GTPase assay

The GTPase activity of purified heterotrimeric G_q was measured in vitro using the GTPase Glo Assay (Promega) with the following modifications. The final reaction consisted of 20 mM HEPES (pH 7.4), 100 mM NaCl, 10 mM MgCl₂, and 1 mg ml⁻¹ BSA. HDL-M2Rpp (12.5 nM) was pre-incubated with iperoxo (10 μM), purified wild-type βarr1-MC-393 or βarr1-MC-393(3×D) (1 μM), and Fab30 (1 μM) for 15 min at room temperature. G-protein (250 nM) and GTP (2.5 μM) were subsequently added, and reactions proceeded for 1 h at room temperature before the addition of GTPase Glo reagent and ADP, as described in the manufacturer's protocol. Luminescence was measured on a CLARIOstar plate reader (BMG Labtech).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic coordinates of the M2R–βarr1 structure have been deposited in the Protein Data Bank under accession number 6UIN. The electron microscopy maps of M2R–βarr1–Fab30(scFv) and interface M2R–βarr1–Fab30(scFv) have been deposited in the Electron Microscopy Data Bank with accession codes EMD-20612 and EMD-20948, respectively.

32. Kruse, A. C. et al. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
33. Nobles, K. N., Guan, Z., Xiao, K., Oas, T. G. & Lefkowitz, R. J. The active conformation of beta-arrestin1: direct evidence for the phosphate sensor in the N-domain and conformational differences in the active states of beta-arrestins1 and -2. *J. Biol. Chem.* **282**, 21370–21381 (2007).
34. Zheng, S. Q. et al. MotionCorr2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e24166 (2018).
36. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
37. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
38. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
39. Adams, P. D. et al. The PHENIX software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
41. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
42. Dolinsky, T. J. et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
43. Tang, C. L., Alexov, E., Pyle, A. M. & Honig, B. Calculation of pK_s in RNA: on the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol.* **366**, 1475–1496 (2007).
44. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
45. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
46. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
47. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
48. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLSAA force field. *J. Chem. Theory Comput.* **11**, 3499–3509 (2015).
49. Dodda, L. S., Cabeza de Vaca, I., Tirado-Rives, J. & Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **45**, W331–W336 (2017).
50. Kulig, W., Pasenkiewicz-Gierula, M. & Róg, T. Topologies, structures and parameter files for lipid simulations in GROMACS with the OPLS-aa force field: DPPC, POPC, DOPC, PEPC, and cholesterol. *Data Brief* **5**, 333–336 (2015).
51. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* (1996).
52. Hirsch, J. A., Schubert, C., Gurevich, V. V. & Sigler, P. B. The 2.8 Å crystal structure of visual arrestin: a model for arrestin's regulation. *Cell* **97**, 257–269 (1999).
53. Luttrell, L. M. et al. Manifold roles of β-arrestins in GPCR signaling elucidated with siRNA and CRISPR/Cas9. *Sci. Signal.* **11**, eaat7650 (2018).
54. O'Hayre, M. et al. Genetic evidence that β-arrestins are dispensable for the initiation of β₂-adrenergic receptor signaling to ERK. *Sci. Signal.* **10**, eaal3395 (2017).

Acknowledgements We thank G. Hodgson, J. Taylor, Q. Lennon and V. Brennan for administrative assistance. Financial support was provided by the National Institutes of Health (grants R01HL16037 to R.J.L. and R01NS092695 to G.S.) and the Mathers Foundation (G.S.). R.J.L. is a Howard Hughes Medical Institute investigator. A.L.W.K. is a Howard Hughes Medical Institute medical research fellow. We thank A. Masoudi for assistance in structural analysis of the M2R–βarr1 complex, S. Zheng for assistance in initially screening M2R–βarr1 complexes by negative-stain electron microscopy, and A. Inoue for βarr1/2-null HEK293 cells.

Author contributions D.P.S. and L.M.W. initiated the project, cloned proteins and optimized M2R expression. D.P.S., A.L.W.K. and W.D.C. expressed and purified components of the HDL-M2R–βarr1 complex. A.L.W.K. and D.P.S. formed and purified HDL-M2R–βarr1 complexes. A.L.W.K. and D.P.S. performed radioligand binding, GTPase Glo, and biman experiments. D.P.S. and L.M.W. conducted cellular assays. H.H. prepared cryo-EM grids, screened conditions, collected images, processed data and reconstructed the final map. H.H. and M.J.R. built and refined the model. M.J.R. performed, analysed and interpreted molecular dynamics simulations with input from N.R.L., H.H., M.J.R. and D.P.S. analysed the structure. D.P.S., L.M.W., G.S., H.H., M.J.R. and R.J.L. prepared the manuscript with input from all authors. R.J.L. and G.S. supervised the project.

Competing interests The authors declare no competing interests.

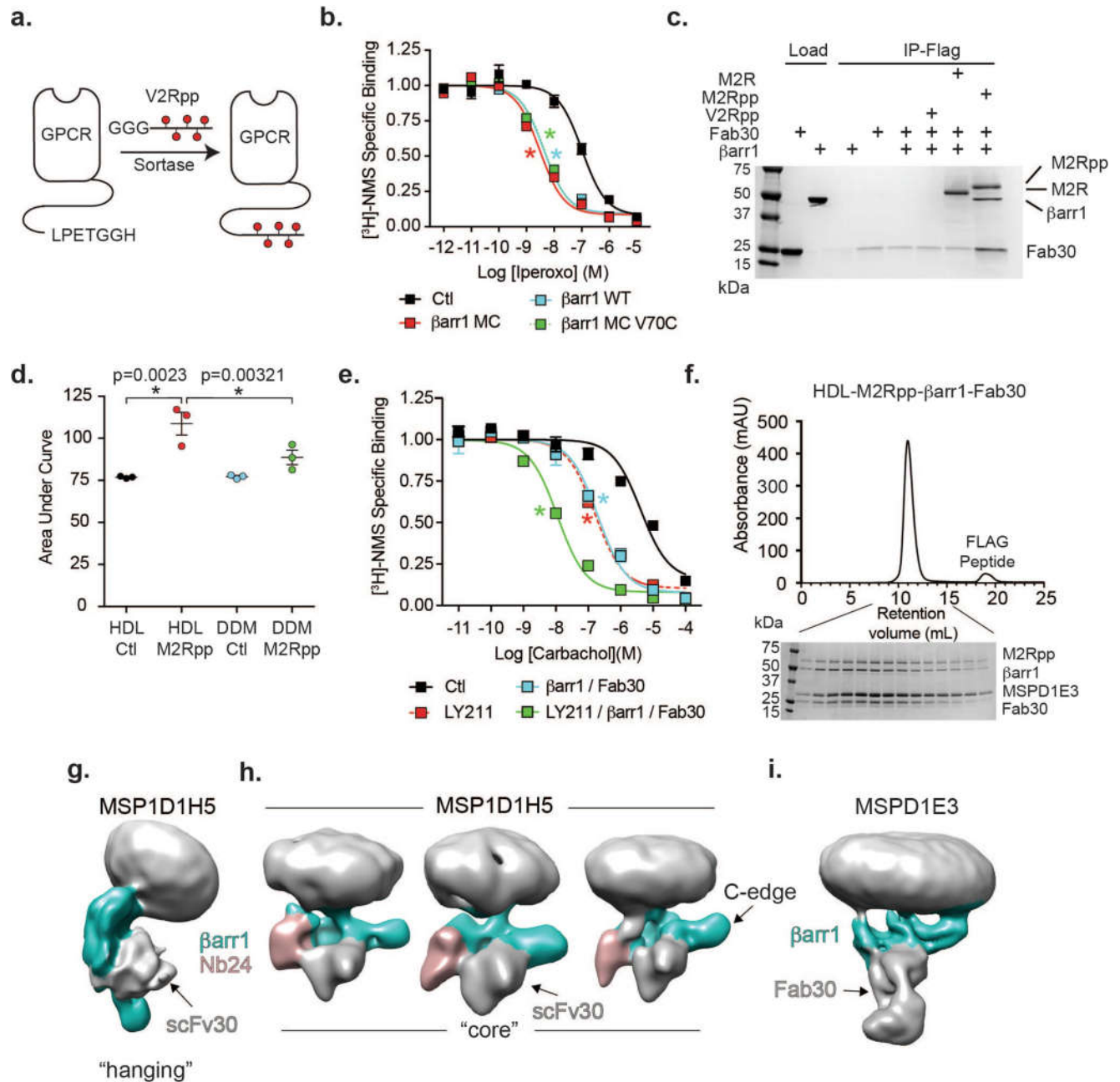
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1954-0>.

Correspondence and requests for materials should be addressed to R.J.L. or G.S.

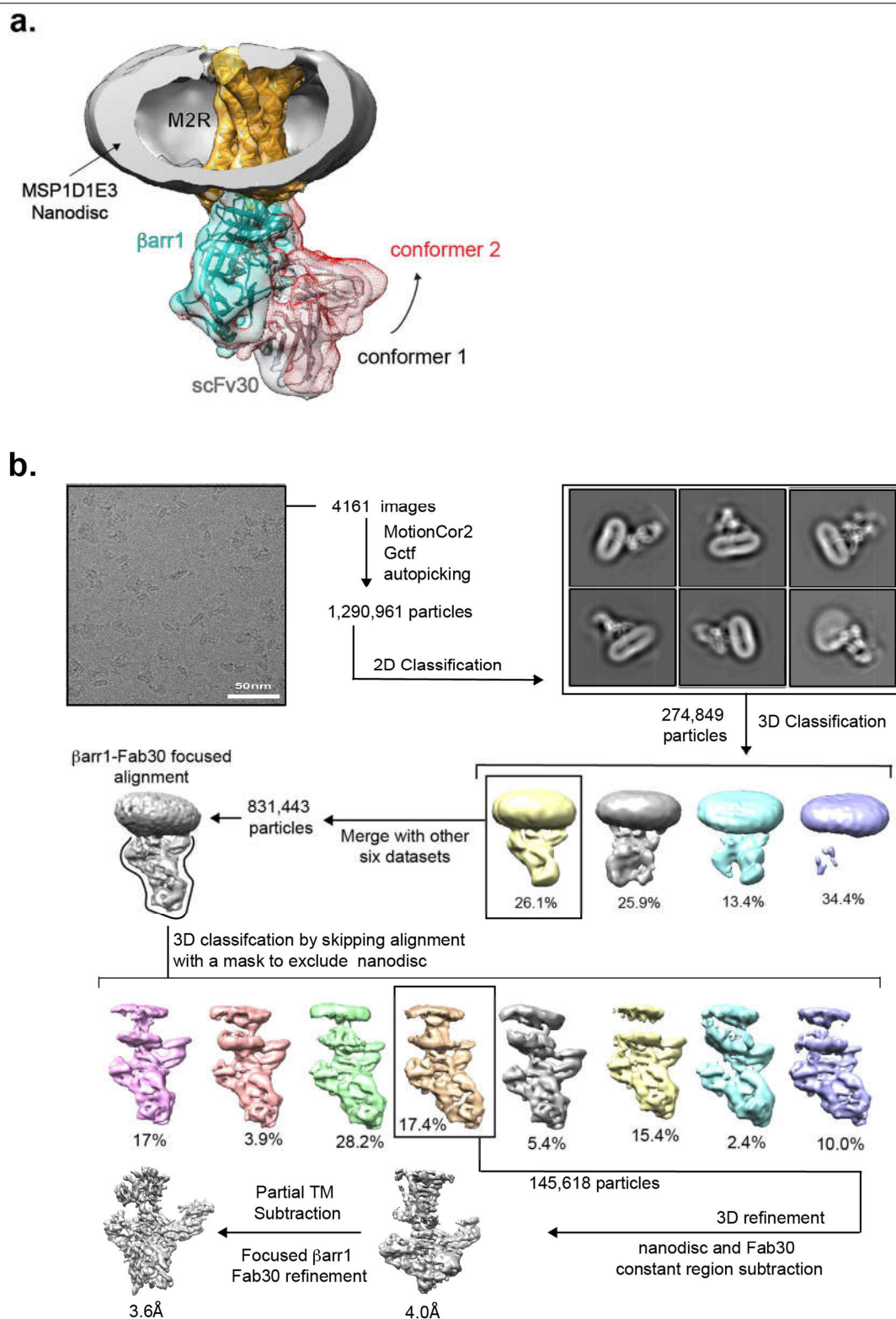
Peer review information Nature thanks Oliver Clarke, Lei Shi, John Tesmer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

**Extended Data Fig. 1 | Characterization of the M2R- β arr1 complex.**

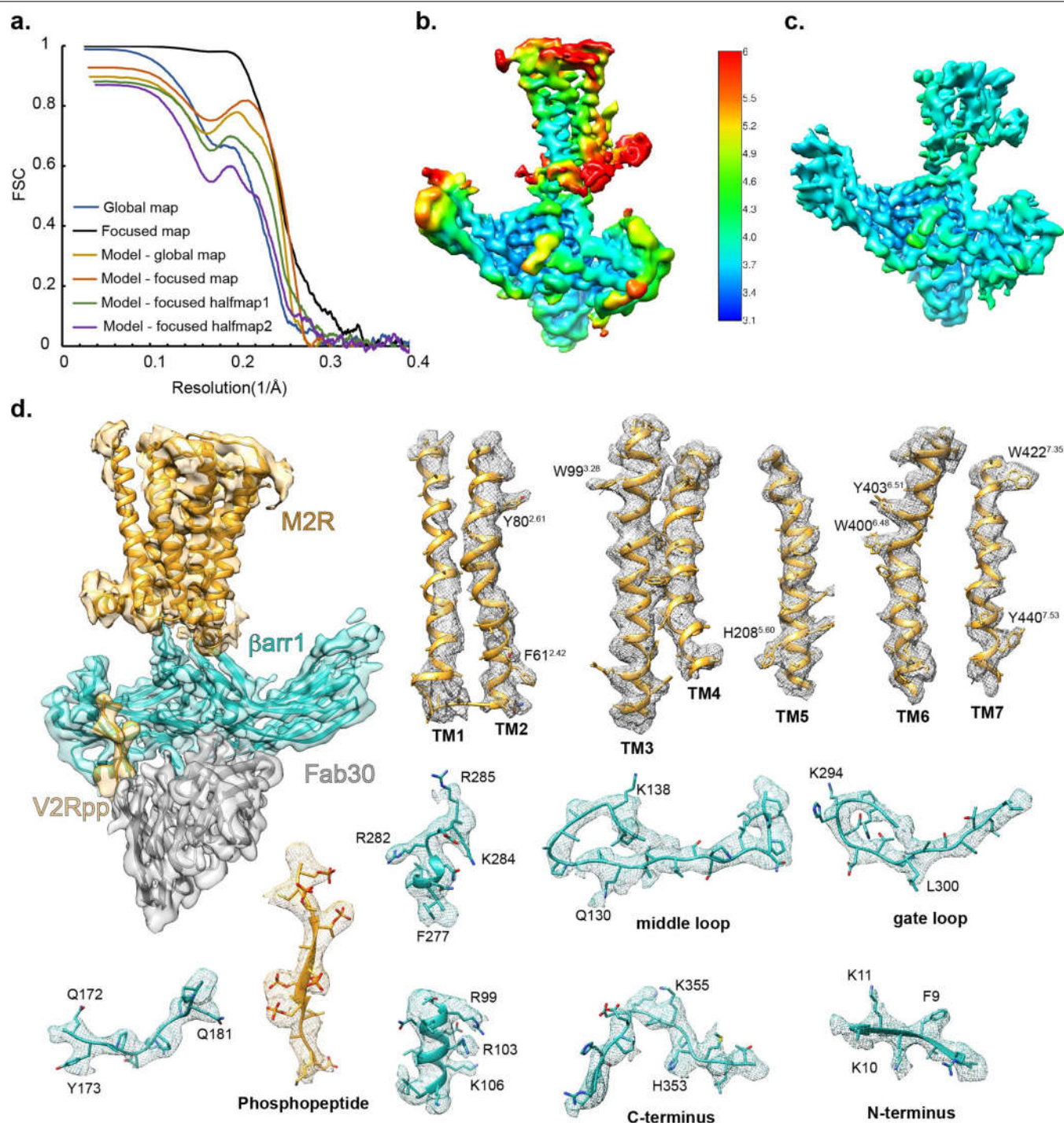
a, Schematic showing sortase-mediated ligation of GGG-V2Rpp onto GPCRs containing a C-terminal sortase consensus sequence (LPETGGH). **b,** Competition radioligand binding experiments using [3 H]-NMS to measure the affinity of iperexo for HDL-M2Rpp in the absence (control; Ctl) ($\log IC_{50} -6.98 \pm 0.07$) or presence of β arr1 ($\log IC_{50} -8.38 \pm 0.07$), β arr1-minimal cysteine (MC) ($\log IC_{50} -8.52 \pm 0.07$) and the mutant β arr1(V70C) ($\log IC_{50} -8.34 \pm 0.06$). **c,** Co-immunoprecipitation (IP) of β arr1 and Fab30 in the presence and absence of DDM-Flag-M2Rpp and DDM-Flag-M2R. Data are representative of three independent experiments. **d,** Statistical analysis of bimane fluorescence data from Fig. 1c. Data are the mean of three independent experiments with error bars representing s.e.m. * indicates significance for the indicated comparison, determined by one-way ANOVA. **e,** Competition radioligand binding experiments using [3 H]-NMS to measure the affinity of the agonist carbachol for

HDL-M2Rpp in the absence (Ctl; $\log IC_{50} -5.38 \pm 0.12$) and presence of LY2119620 (LY211) ($\log IC_{50} -6.7 \pm 0.10$), β arr1 and Fab30 ($\log IC_{50} -6.8 \pm 0.07$) or in combination ($\log IC_{50} -7.95 \pm 0.06$). **f,** Size-exclusion chromatography of the final MSP1D1E3-M2Rpp- β arr1-Fab30 complex and SDS-PAGE analysis of peak fractions. **g, h,** Low resolution cryo-EM analysis of M2Rpp- β arr1-Nb24-scFv30 complex in MSP1D1H5 nanodiscs showing β arr1 in a 'hanging' conformation (g) or 'core' conformations with 'rocking' relative to the nanodisc density (h). **i,** Low-resolution cryo-EM map of the M2Rpp- β arr1-Fab30 complex in the larger MSP1D1E3 nanodiscs shows β arr1 in the 'core' conformation involving an additional interaction of the C domain with the lipid bilayer. All β arr1 variants are truncated at amino acid 393. Radioligand binding experiments are the means of three independent experiments with error bars representing s.e.m. * indicates significance compared to control ($P < 0.0001$, one-way ANOVA).



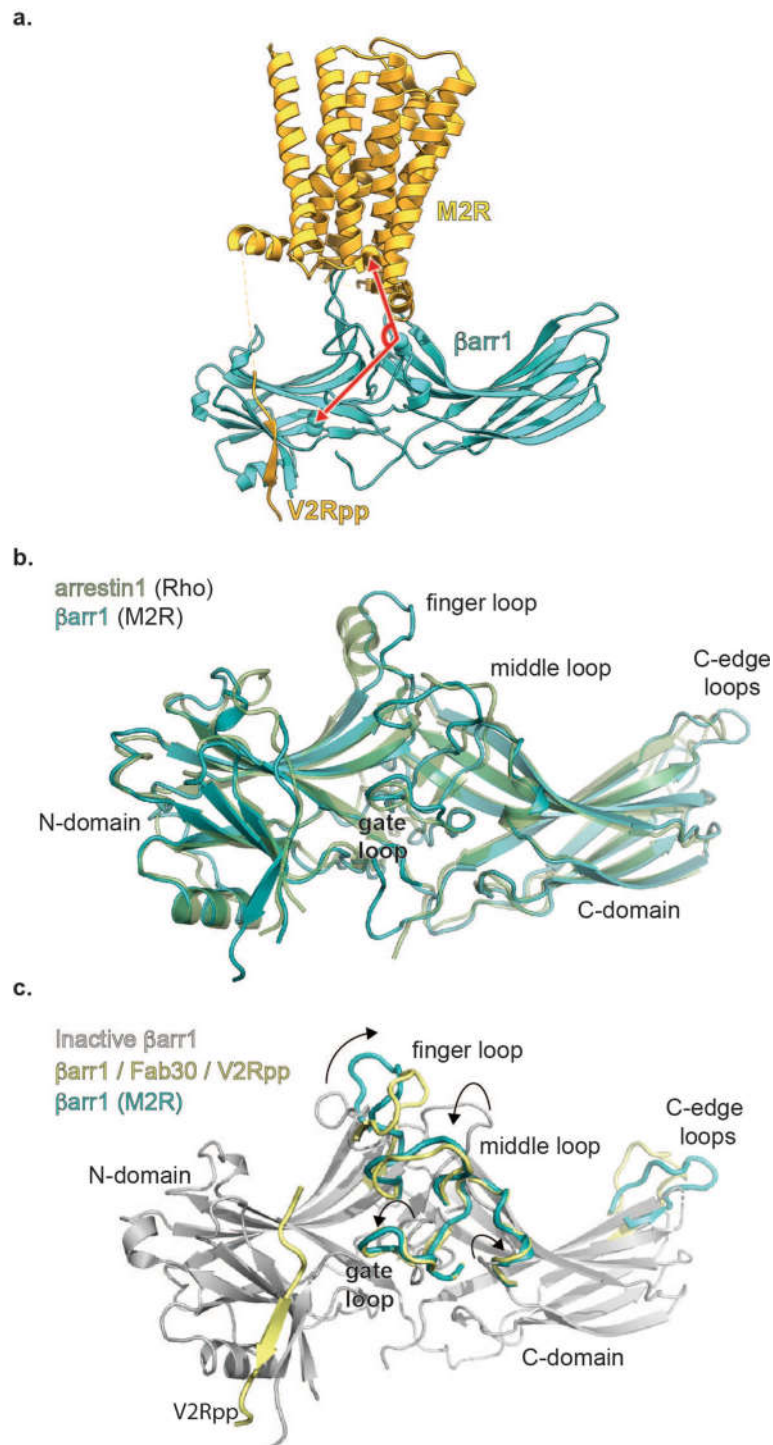
Extended Data Fig. 2 | M2Rpp-βarr1 complex conformers and cryo-EM reconstruction workflow. **a.** Conformational variability of the M2Rpp-βarr1 complex. Overlay of two low resolution reconstructions aligned on the 7TM portion reveal variability in the angle of the βarr1-Fab30 segment relative to the receptor. One map is shown as mesh and the other as solid surface. The constant domains of Fab30 have been masked out in the reconstructions. For

clarity, only one receptor-nanodisc density is shown. **b.** Flow chart of cryo-EM data processing towards high-resolution reconstructions. All eight-particle classes show engagement of βarr1 with the lipid nanodisc, but only one class (17.4% of particles) could be further processed to high resolution. A final focused refinement yielded a 3.6 Å structure providing insights into the binding interface and orientation of βarr1 relative to M2Rpp.



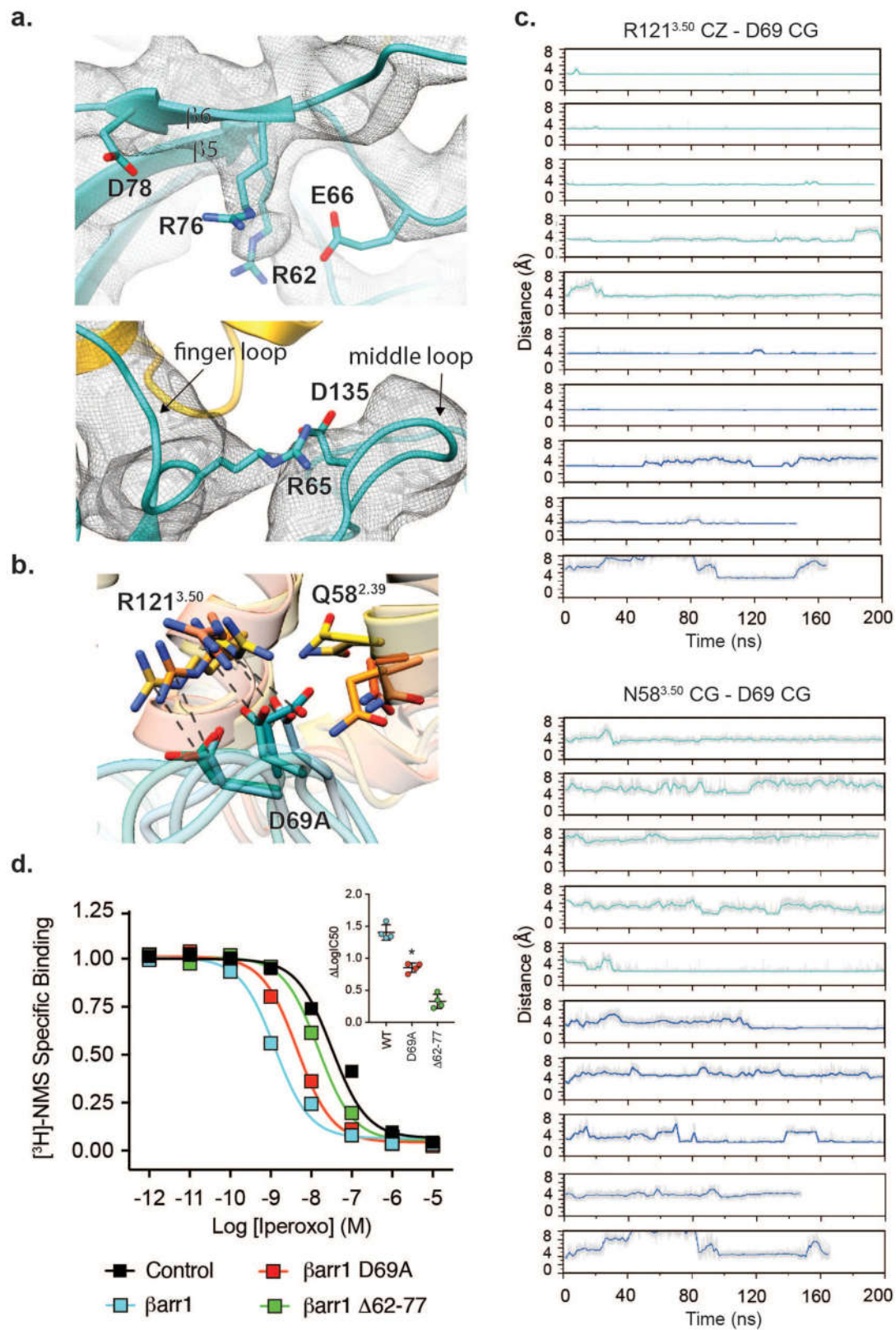
Extended Data Fig. 3 | Map resolution and model validation. **a**, 'Gold standard' FSC plots of the reconstruction of M2Rpp-βarr1-Fab30(scFv) (global map, black) and focused refinement reconstruction for interface M2Rpp-βarr1-Fab30(scFv) (focused map, blue). The red and brown curves represent the correlation between model and map for the global map and the focused map, respectively. Refined model validation was also performed by calculating the correlation between model and focused half map 1 (green), and the

correlation between the randomly displaced model and focused half map 2 (purple). **b, c**, Local resolution estimation map for M2R-βarr1-Fab30(scFv) and focused map of half-M2Rpp-βarr1-Fab30(scFv). **d**, Overall model fit to the cryo-EM map. The 7TM portion (orange) comes from the full M2Rpp-βarr1-Fab30(scFv) map contoured at $\sigma=7.5$, whereas the βarr1-scFv portion comes from the focused refinement of Interface M2Rpp-βarr1-Fab30(scFv) contoured at $\sigma=5.2$.



Extended Data Fig. 4 | Structural comparisons of arrestins. **a.** The tilt of β -arrestin relative to the M2Rpp was measured by the angle (red arrow) between V37 (β arr1), I317 (β arr1) and R121^{3.50} (M2R). **b.** Superposition of β arr1 in the M2Rpp- β arr1 complex and visual arrestin in the crystal structure of Rho-

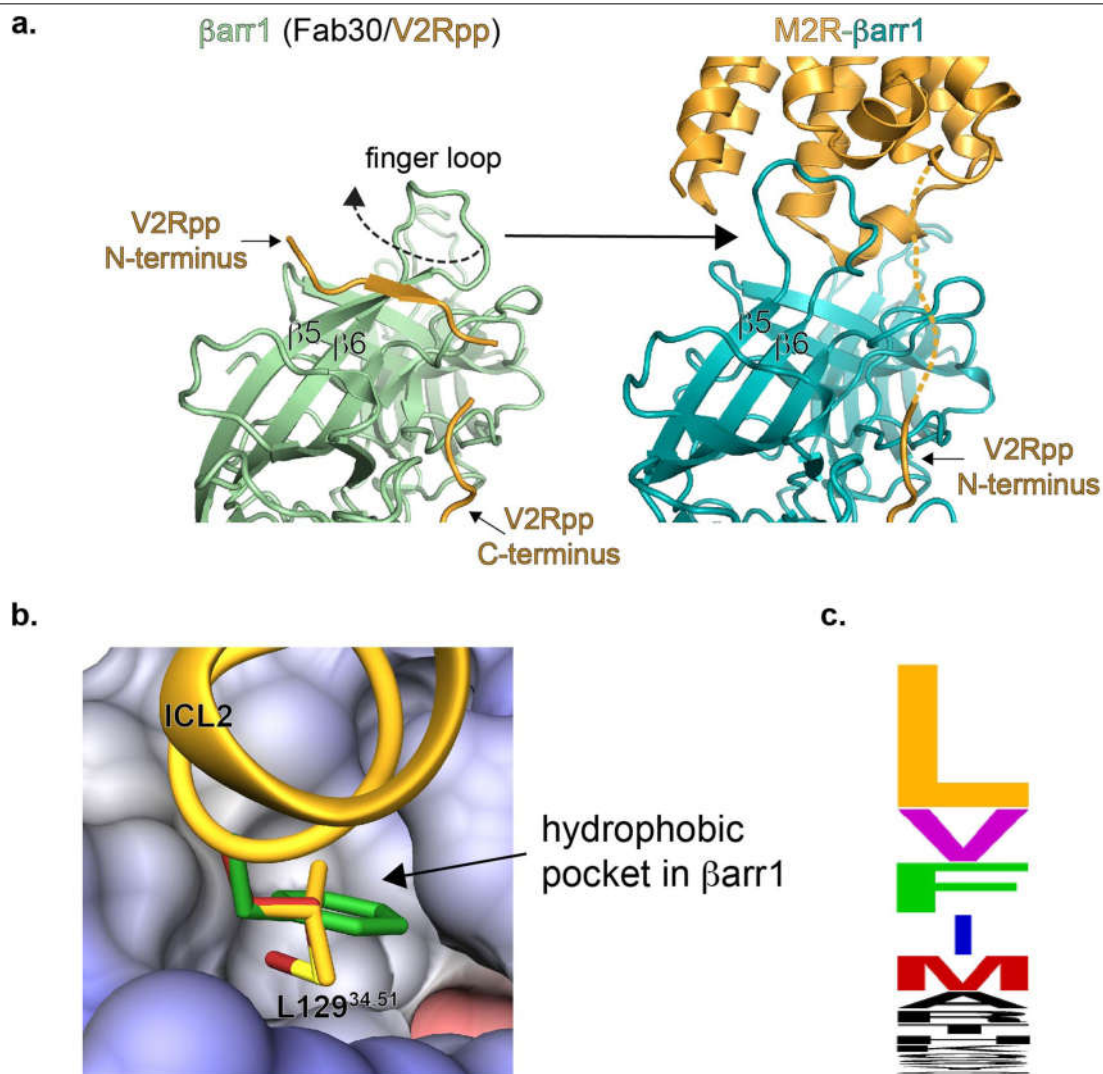
Arr1 (PDB: 5W0P). **c.** Superposition of β arr1 in the inactive state (grey, PDB: 1G4M), in the phosphopeptide-activated state (green, PDB: 4JQI) and in the receptor-bound active state (gold, M2Rpp- β arr1). The arrows indicate rearrangements of central crest loops.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Computational and experimental examination of interactions between the finger loop and M2Rpp. **a**, Depiction of E61 of the finger loop and probable interaction partners R57 and R71 (top) and the R60 of the finger loop and D130 of the middle loop (bottom) from PHENIX/OPLS3 refinement in the cryo-EM map. The mesh depicts the 3.6 Å cryo-EM map contoured at $\sigma = 4.0$ with a masked 3.0 Å zone around the atoms depicted (top) or contoured at $\sigma = 3.0$ with a masked 3.0 Å zone around the atoms depicted (bottom). **b**, An overlay of the last frame from 200 ns of simulation for five of the molecular dynamics simulation trajectories. D69 of arrestin is depicted as bonds in addition to R^{3.50} and N^{2.39} of M2Rpp. Lines connect D69 and R^{3.50} from the same snapshot. **c**, Plots of the R^{3.50} zeta carbon–D69 gamma carbon distance (top) and N^{2.39} gamma carbon–D69 gamma carbon distance (bottom)

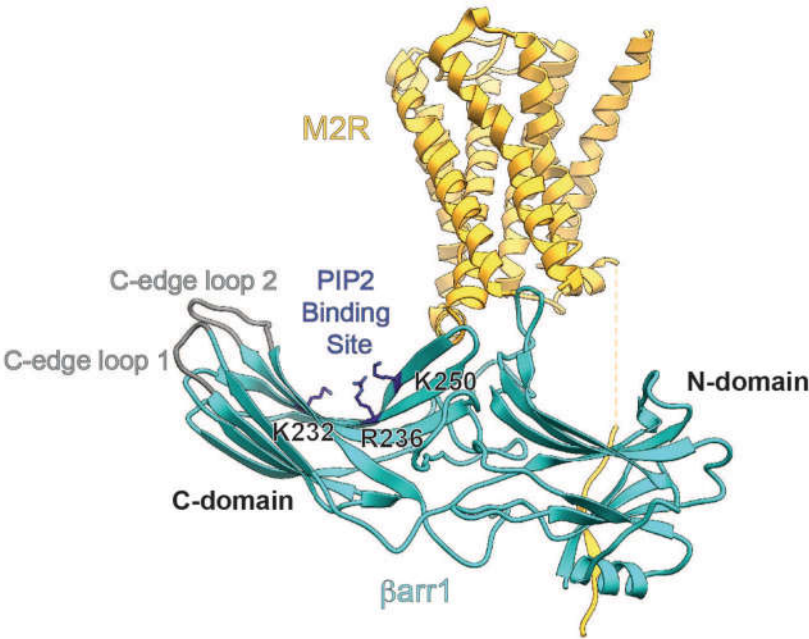
over the course of the molecular dynamics simulations performed in this work. Grey lines correspond to raw data, whereas coloured lines correspond to a 1-ns sliding average. Teal traces correspond to simulations with membrane, and blue traces correspond to simulations with a small nanodisc. **d**, Competition radioligand binding experiments using [³H]NMS to measure the affinity of the agonist iperovo for HDL-M2Rpp in the absence and presence of wild-type β arr1, β arr1(D69A) and the β arr1(Δ 62–77) mutant that lacks the finger loop. Data are the mean of four independent experiments with error bars representing s.e.m. Inset, difference in logIC₅₀ between β arr1 variants and control (no β arr1). *denotes significance compared to wild-type β arr1 ($P < 0.0001$, one-way ANOVA).



Extended Data Fig. 6 | Finger loop rearrangements and ICL2. **a**, Detail of crystal structure of activated β arr1 with phosphopeptide (left, PDB: 4JQI) showing the N-terminal portion of V2Rpp bound on β -strands 5 and 6, thereby twisting the finger-loop fold. Unbinding of the N-terminal portion of V2Rpp is required for the finger loop to adopt the observed conformation in the cryo-EM structure of M2Rpp- β arr1 (right). The arrow shows the direction of finger-loop

untwisting. **b**, Expanded view of L129 (orange) in the cleft of β arr1 overlaid with a modelled phenylalanine (green) and methionine (red) at the same position. **c**, Plot of the frequency of specific amino acids occurring in the second position of ICL2 for β arr-binding class A GPCRs. The size of the one-letter code is correlated to the frequency with which that residue occurs at that position.

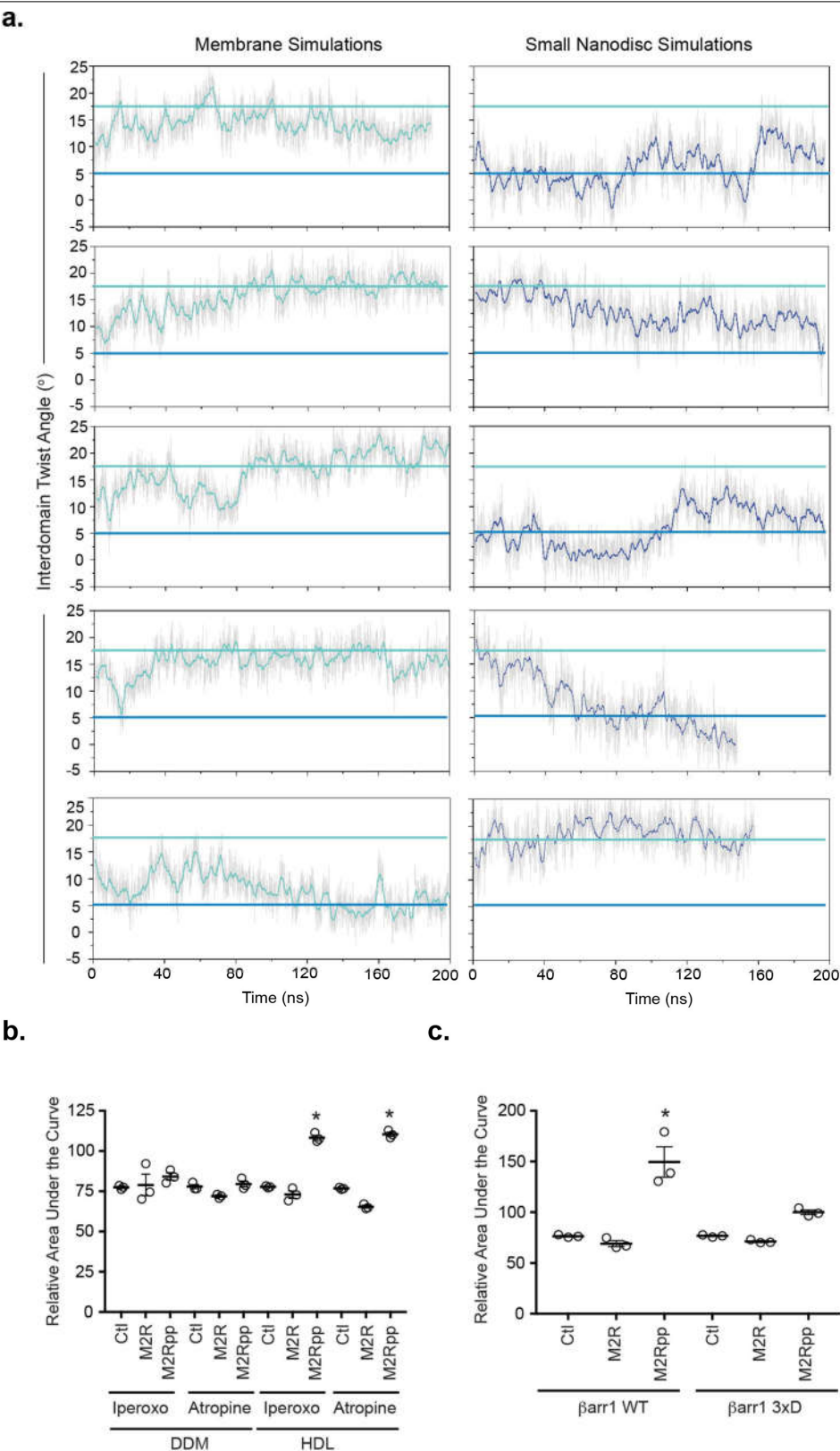
a.



b.

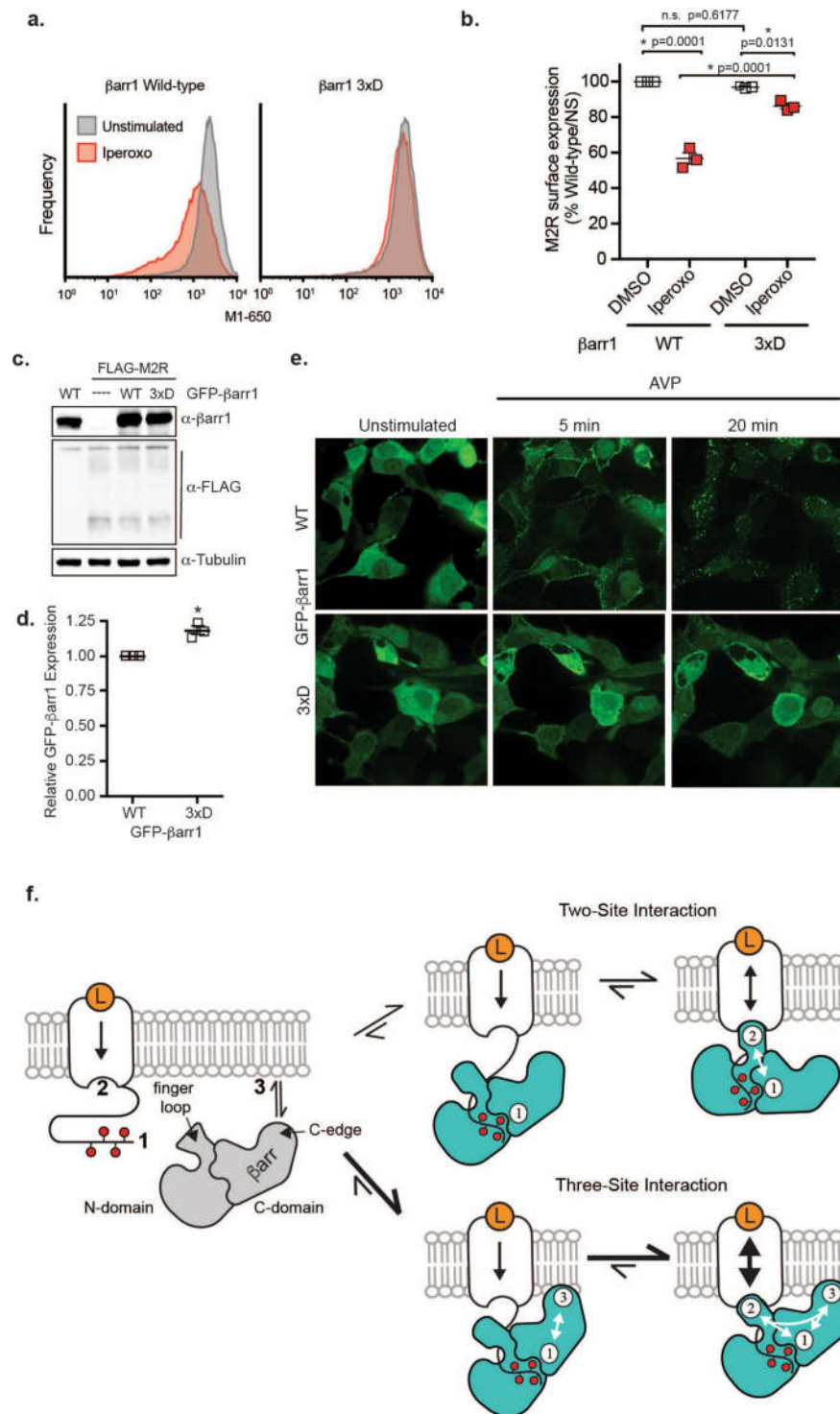
	C-edge loop 1	PIP2 binding	PIP2 binding	C-edge loop 2
β-arrestin 1, isoform A	188 RQFLMSDKPLHL	230 KIKISVRQY	248 QYKCP	327 LVVSRGGLGLASSDV
β-arrestin 1, isoform B	188 RQFLMSDKPLHL	230 KIKISVRQY	248 QYKCP	327 LVVSRGG-----DV
β-arrestin 2	189 RHFLMSDRSLHL	231 KIKISVRQY	249 QYKCP	328 LVVSRGG-----DV
X-arrestin	185 RRFLLSAQPLQL	227 KIKISVDQI	245 KYTKT	324 LMVSCGGILGLTASDV
visual arrestin	198 WQFFMSDKPLHL	240 KIKAFVFEQV	258 YVVKP	337 LTVS--GFLGELTSSEV

Extended Data Fig. 7 | Conservation of lipid-binding regions in the arrestin family. **a**, Ribbon model of the M2Rpp-βarr1 complex depicting the location of C-edge loops and PtdIns(4,5)P₂ (PIP2)-interacting residues. **b**, Sequence alignment of the arrestin family, showing differential conservation of the C-edge loops and the PtdIns(4,5)P₂-binding motif.



Extended Data Fig. 8 | Effects of the lipid membrane on β arr1 coupling to M2Rpp. **a.** Atomistic simulations of the M₂Rpp- β arr1 complex in a membrane bilayer or a small nanodisc. Time courses are provided for the calculated interdomain twist angle for each replicate in membrane bilayer (left) or a small nanodisc (right). Raw data are provided in grey and a 1-ns rolling average is provided in teal for the membrane simulations and blue for the small nanodisc simulations. Horizontal teal and blue lines correspond to an active-like and

inactive-like interdomain twist angle, respectively. **b.** Statistical analysis of the data from Fig. 4c; data represent means of three independent experiments with error bars representing s.e.m. * $P < 0.0001$, one-way ANOVA; indicating significance from control within the same subset. **c.** Statistical analysis of data from Fig. 4d; data represent the mean and s.e.m. of three independent experiments. * $P < 0.0001$, one-way ANOVA; denotes significance compared to control.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | β arr1 C-edge–lipid interaction facilitates M2R

internalization. **a**, Flow cytometry analysis of β arr1/2-null cells transiently transfected with Flag–M2R and GFP– β arr(WT) or GFP– β arr1(3 \times D). Cells were treated with vehicle or iperoxo for 30 min and subsequently stained with Alexa Fluor-650-labelled anti-Flag M1 antibody. GFP⁺ singlet cells were gated for analysis. Results are representative of data from three independent experiments. **b**, Quantitation of Flag–M2R surface staining by flow cytometry, as described above. Alexa Fluor-650-labelled anti-Flag M1 staining was normalized to the mean fluorescence of unstimulated cells expressing wild-type β -arrestin1–GFP in each experiment. These data were used to calculate the percentage of receptor internalized in Fig. 5c. Data represent the mean and standard error from three independent experiments and asterisks indicate statistical significance (one-way ANOVA). n.s., not significant. **c**, Expression of GFP– β arr1(WT) or GFP– β arr1(3 \times D) and Flag–M2R in β arr1/2-null HEK293 cells as assessed by SDS–PAGE and western blot analysis. Tubulin was used as a loading control. Data are representative of three independent experiments.

d, Quantification of GFP– β arr1(WT) or GFP– β arr1(3 \times D) expression by flow cytometry using β arr1/2-null HEK293 cells. Data represent the mean and standard error from three independent experiments; * P = 0.0034 (two-sided unpaired t -test). **e**, Localization of GFP– β arr1(WT) or β arr1(3 \times D) in Flag–vasopressin-2-receptor overexpressing HEK293 cells treated with arginine vasopressin peptide (AVP) for the indicated time. Data are representative of three independent experiments. **f**, Three-site interaction network of GPCR– β -arrestin binding. In the classic two-site interaction model, conformational changes in β -arrestin induced by binding to phosphorylated receptor (1) leads to transmembrane receptor core coupling (2) to sterically block G protein binding. Our findings suggest an expanded model including interaction of the C domain of β -arrestin with the lipid bilayer (3) because it synergistically enhances the interaction of β -arrestin with the phosphorylated receptor tail/loops and transmembrane core. Vertical arrows in the receptor represent direction and strength of cooperativity between the extracellular orthosteric ligand-binding and intracellular transducer-binding sites.

Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statistics

	M2R- β arr1- Fab30(scFv) (EMDB-20612) (PDB 6U1N)	InterfaceM2R- β arr1- Fab30(scFv) (EMDB- 20948)
Data collection and processing		
Magnification	47,169	47,169
Voltage (kV)	300	300
Electron exposure (e-/Å ²)	50	50
Defocus range (μm)	1.2-2.2	1.2-2.2
Pixel size (Å)	1.06	1.06
Symmetry imposed	C1	C1
Initial particle images (no.)	11.7million	11.7million
Final particle images (no.)	145,618	145,618
Map resolution (Å)	4Å	3.6Å
FSC threshold	0.143	0.143
Map resolution range (Å)	3.6-6	3.2-5
Refinement		
Initial model used (PDB code)	6OIK 4JQI	
Model resolution (Å)	4.03	
FSC threshold	0.5	
Map sharpening <i>B</i> factor (Å ²)	150	
Model composition		
Non-hydrogen atoms	5863	
Protein atoms	5834	
Ligands	29	
<i>B</i> factors (Å ²)		
Protein	164.85	
Ligand	292.74	
R.m.s. deviations		
Bond lengths (Å)	0.005	
Bond angles (°)	0.682	
Validation		
MolProbity score	2.2	
Clashscore	18.0	
Poor rotamers (%)	1.0	
Ramachandran plot		
Favored (%)	91.6	
Allowed (%)	8.4	
Disallowed (%)	0.0	

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
*Give *P* values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

SerialEM 3.6 software was used for automated cryoEM data collection.

Data analysis

Data from radioligand binding, bimane fluorescence, and GTPase Glo assay were analyzed using GraphPad Prism v7. Flow cytometry data were analyzed using Prosort v.1.5.0.15 and FlowJo 10. CryoEM image processing and 3D reconstructions conducted using MotionCor2 1.3.0, Gctf 1.06, relion3.0 and cisTEM 1.0.0. Bsoft 2.0.5 was used for local resolution determination and Phenix 1.15 for map fsc curve calculation and model validation. Coot 0.8.9.1 was used for model building. Molecular dynamics experiments were run with NAMD v2.13 and analyzed with VMD v1.9.3. Chimera v1.12 was used for electrostatic analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Atomic coordinates have been deposited in the Protein Data Bank (PDB) under accession number 6U1N and the cryoEM density maps in EMDB under number 20612 and 20948

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Since no inferences were determined for a population based on a sample, no sample size determination was needed in this study,
Data exclusions	No data were excluded from analysis
Replication	A minimum of three independent experiments were conducted for radioligand binding, flow cytometry, bimanual fluorescence, confocal microscopy, and GTPase Glo assays. All attempts at replication were successful with no data exclusions.
Randomization	The independent variables of the experimental design were well controlled and did not require randomization
Blinding	Experimental results were all quantitative and did not require any subjective analysis, thus no experiments were performed with blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a Involved in the study

☐ ☒ Antibodies

☐ ☒ Eukaryotic cell lines

☒ ☐ Palaeontology

☒ ☐ Animals and other organisms

☒ ☐ Human research participants

☒ ☐ Clinical data

n/a Involved in the study

☒ ☐ ChIP-seq

☐ ☒ Flow cytometry

☒ ☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-FLAG M1 used for receptor purification, flow cytometry (120nM), and confocal microscopy (30nM) was purified from ATCC hybridoma line HB-9259 (Clone 4E11). Anti-FLAG-M2-HRP (Sigma; A8592) and Tubulin (Sigma; T7816) were used in western blots at concentrations of 1:1000 and 1:5000, respectively.
Validation	Certificate of analysis of FLAG-M1 is available from ATCC website and was further validated by pull-down assays and flow cytometry. Specificity of FLAG-M2-HRP was determined by including control cell lysate not expressing FLAG-tagged receptor. Tubulin antibody was validated by western blot analysis and compared to that illustrated on the Sigma specification sheet

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Expi293F cells: ThermoFisher. The derivative tetracycline-inducible Expi293F cell line used for M2R expression was reported in Staus et. al. Proc Natl. Acad. Sci. USA 115, 3834-3839 (2018). Beta-arrestin1/2 knockout HEK293 cell lines (Dr. Asuka Inoue, Tohoku University) were characterized and reported in Grundmann et. al. Nat Commun. 2018;9(1):341.
Authentication	No additional authentication of cell lines was performed
Mycoplasma contamination	Cells were not tested for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

beta-Arrestin1/2 null HEK293 cells (courtesy of A. Inoue) overexpressing GFP-beta-arrestin1 constructs were stained with AlexaFluor650-M1 (anti-FLAG) antibody.

Instrument

Bio-Rad S3e Cell Sorter

Software

Data were collected using ProSort software and analyzed in FlowJo 10.

Cell population abundance

N/A, cells were not sorted.

Gating strategy

N/A

- ☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Structure of the neurotensin receptor 1 in complex with β -arrestin 1

<https://doi.org/10.1038/s41586-020-1953-1>

Received: 12 August 2019

Accepted: 8 January 2020

Published online: 16 January 2020

 Check for updates

Weijiao Huang^{1,7}, Matthieu Masureel^{1,7}, Qianhui Qu^{1,2,7}, John Janetzko^{1,7}, Asuka Inoue³, Hideaki E. Kato^{1,6}, Michael J. Robertson^{1,2}, Khanh C. Nguyen⁴, Jeffrey S. Glenn⁴, Georgios Skiniotis^{1,2,5}✉ & Brian K. Kobilka¹✉

Arrestin proteins bind to active, phosphorylated G-protein-coupled receptors (GPCRs), thereby preventing G-protein coupling, triggering receptor internalization and affecting various downstream signalling pathways^{1,2}. Although there is a wealth of structural information detailing the interactions between GPCRs and G proteins, less is known about how arrestins engage GPCRs. Here we report a cryo-electron microscopy structure of full-length human neurotensin receptor 1 (NTSR1) in complex with truncated human β -arrestin 1 (β arr1(Δ CT)). We find that phosphorylation of NTSR1 is critical for the formation of a stable complex with β arr1(Δ CT), and identify phosphorylated sites in both the third intracellular loop and the C terminus that may promote this interaction. In addition, we observe a phosphatidylinositol-4,5-bisphosphate molecule forming a bridge between the membrane side of NTSR1 transmembrane segments 1 and 4 and the C-lobe of arrestin. Compared with a structure of a rhodopsin–arrestin-1 complex, in our structure arrestin is rotated by approximately 85° relative to the receptor. These findings highlight both conserved aspects and plasticity among arrestin–receptor interactions.

Upon activation, GPCRs signal through G-protein pathways and arrestin pathways to regulate downstream cellular events. Recent studies suggest that drugs that direct signalling through only one of these pathways (known as biased drugs) may have fewer side effects^{3–5}. There are four isoforms of arrestin, of which arrestin-1 (Arr1) and arrestin-4 (Arr4) are known as visual arrestins. Arr1 is expressed in retinal rods and is important in turning off activated rhodopsin, whereas Arr4 is expressed in cones and deactivates colour opsins. The non-visual arrestins—arrestin-2 and arrestin-3, also known as β -arrestin 1 (β arr1) and β -arrestin 2 (β arr2)—are important in suppressing G-protein signalling for most other GPCRs. In addition to their role in desensitization, non-visual arrestins regulate GPCR endocytosis through interactions with the clathrin-dependent endocytic machinery, and can act as scaffolds for various other cytosolic signalling molecules downstream of GPCR activation, such as Src family tyrosine kinases and MAP kinases^{1,2,6,7}. Owing to recent breakthroughs in structural biology, structures of family A GPCR–G-protein complexes have been resolved for the three major classes of G proteins (G_s , $G_{i/o}$ and $G_{q/11}$)^{8–13}; however, the only available structure of a GPCR–arrestin complex is that of rhodopsin–Arr1¹⁴. Despite the structural similarities between rhodopsin and other family A GPCRs, there are fundamental differences in their activation kinetics, phosphorylation and arrestin coupling¹⁵. To better understand the molecular details of the arrestin pathway, and how arrestins can interact with a diverse array of GPCRs, we initiated efforts to obtain structures of non-rhodopsin GPCR– β -arrestin complexes.

β arr1 was first identified as a regulator of β_2 adrenergic receptor (β_2 AR) signalling¹⁶. Although the β_2 AR is one of the most extensively studied model systems for family A GPCRs, initial efforts to form a stable β_2 AR– β arr1 complex for structural studies were not successful. To identify a more stable GPCR– β -arrestin complex, we devised a fluorescence-based screen in which various candidate family A receptors were expressed in HEK293 cells and evaluated for their ability to couple to activated β arr1 in a manner independent of the phosphorylation status of the receptor (Extended Data Fig. 1). This screen revealed that the neurotensin receptor 1 (NTSR1) complexed with β arr1 in vitro, and showed promise as a candidate for structural studies.

NTSR1 mediates responses to neurotensin (NTS) and neuromedin N, both of which are derived from the same precursor peptide¹⁷. NTS regulates a broad spectrum of physiological processes¹⁸ including blood pressure, ileum contraction or relaxation, analgesia and hypothermia. Although there are three subtypes of neurotensin receptor, most of the physiological responses of NTS are mediated by NTSR1. NTSR1 is a promiscuous GPCR, which couples to G_s , $G_{q/11}$, $G_{i/o}$ and $G_{12/13}$ ¹⁸. Both active- and inactive-state structures of NTSR1 have been determined by crystallography^{19–21}, and recently structures of the NTSR1– G_i complex have been determined by cryo-electron microscopy (cryo-EM)⁹. Notably, the cryo-EM analysis revealed two distinct conformations of the NTSR1– G_i complex, termed the canonical state (C-state) and the non-canonical state (NC-state). The C-state is similar to other GPCR–G-protein complexes, whereas the NC-state may represent

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. ²Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. ³Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan. ⁴Departments of Medicine and Microbiology & Immunology, Stanford University, Stanford, CA, USA. ⁵Department of Photon Science, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA, USA. ⁶Present address: Komaba Institute for Science, The University of Tokyo, Tokyo, Japan. ⁷These authors contributed equally: Weijiao Huang, Matthieu Masureel, Qianhui Qu, John Janetzko. ✉e-mail: yjorgo@stanford.edu; kobilka@stanford.edu

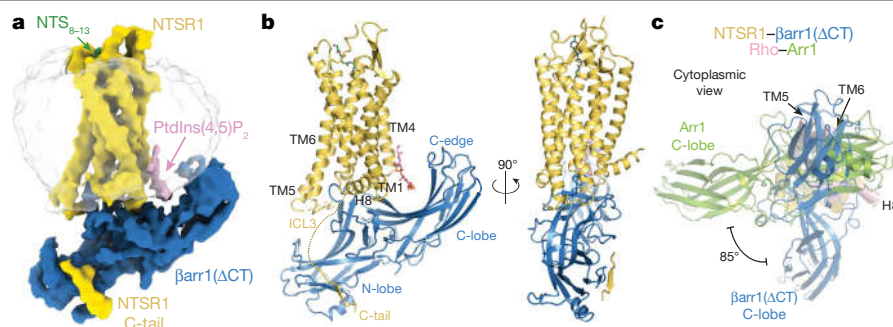


Fig. 1 | The NTSR1-βarr1(ΔCT) complex features a distinct arrestin orientation and unique interaction sites. a, b, Cryo-EM map (a) and model (b) of phosphorylated NTSR1 (yellow) bound to the peptide agonist NTS₈₋₁₃ (green) in complex with βarr1(ΔCT) (blue). Interactions between the receptor and βarr1(ΔCT) are mediated by the phosphorylated C terminus of NTSR1, the C-terminal part of the phosphorylated ICL3, the receptor core and a PtdIns(4,5)P₂ molecule (pink). The C-edge of βarr1(ΔCT) is in contact with the detergent

an intermediate along the activation pathway⁹. Recent studies have shown that an arrestin-biased positive allosteric modulator for NTSR1 suppressed self-administration of methamphetamine in rats⁵, providing further incentive to elucidate the structure of the NTSR1-βarr1 complex.

Preparation of the NTSR1-βarr1 complex

Given the importance of receptor phosphorylation for arrestin coupling, we chose to use full-length, native NTSR1 bound to the agonist NTS₈₋₁₃ (amino acids 8–13 of NTS), and phosphorylated the receptor in vitro using GPCR kinase subtype 5 (GRK5) following a protocol established for the β₂ adrenergic receptor²². GPCR kinases (GRKs) facilitate arrestin-mediated desensitization and signalling by phosphorylating the receptor on its intracellular loops and its C terminus, which increases the ability of the receptor to bind and activate arrestins²³. Previous analysis of GRK-mediated phosphorylation of NTSR1 indicated that whereas both GRK2 and GRK5 phosphorylate the C terminus, GRK5 also phosphorylates the third intracellular loop (ICL3)²⁴. The extent of phosphorylation was monitored by analytical ion-exchange chromatography (Extended Data Figs. 2a, b, 3a). The unphosphorylated receptor did not complex to full-length βarr1, whereas GRK5-phosphorylated NTSR1 did (Extended Data Fig. 2c, d). Moreover, pre-activated βarr1-truncated at residue 382 (βarr1(ΔCT))²⁵—enhanced this coupling (Extended Data Fig. 2c, d). We therefore chose to use this construct for subsequent experiments.

Negative-stain electron microscopy images of the purified complex²⁶ showed particles in which arrestin was tightly engaged with the core of the receptor; however, much of the complex dissociated during vitrification. To address this challenge, we treated the purified complex with a heterobifunctional crosslinker, sulfo-succinimidyl 6-(4,4'-azipentanamido)hexanoate (sulfo-LC-SDA), which has an amine-reactive group on one end and a diazirine moiety on the other. After labelling accessible lysine residues, the complex was exposed to ultraviolet light to activate the diazirine and trigger crosslinking, followed by purification through a second round of size-exclusion chromatography (Extended Data Fig. 3a). The sample seemed qualitatively unchanged according to negative-stain electron microscopy images (Extended Data Fig. 3b) and, although we observe a relatively small fraction of crosslinked complex by SDS-PAGE, the photo-crosslinked sample remained intact during sample vitrification and afforded homogeneously dispersed particles that were suitable for cryo-EM. Mass spectrometry was used to analyse the effect of the photo-crosslinker on the

micelle (translucent grey). c, Overlay of the NTSR1-βarr1(ΔCT) structure with the rhodopsin-arrestin-1 structure (with rhodopsin and Arr1 coloured pink and green, respectively), viewed from the cytoplasmic side. The structures were aligned on the basis of the receptor chains (helices shown as cylinders). Compared with βarr1(ΔCT), Arr1 is rotated by approximately 85°. Additional views of this overlay are shown in Extended Data Fig. 7a.

NTSR1-βarr1(ΔCT) complex. We found that, although several lysine residues on the N-lobe of βarr1(ΔCT) as well as in the ICL1 and ICL2 regions of NTSR1 reacted with the crosslinker, they largely resulted in dead-end crosslinks during photo-illumination (Extended Data Fig. 4, Supplementary Tables 1–7). The majority of the observed crosslinks were intramolecular and largely confined to ICL2 of the NTSR1 and the loops of the N-lobe of βarr1(ΔCT). Only a single intermolecular crosslink that was consistent with our structural model was identified—between the cytoplasmic end of receptor transmembrane helix 1 (TM1) and a β-strand of βarr1(ΔCT) (Extended Data Fig. 4c)—although its density is not observed in our map. It is therefore possible that the main effect of crosslinking was to stabilize βarr1(ΔCT) in a native conformation with enhanced affinity for NTSR1, thereby enabling the complex to survive during cryo-EM grid preparation.

Cryo-EM of the NTSR1-βarr1(ΔCT) complex

A large dataset of 18,000 micrographs of the photo-crosslinked NTSR1-βarr1(ΔCT) complex was processed in subsets and combined to yield about 600,000 particles from well-defined three-dimensional (3D) classes, which were reconstructed to an overall 4.5 Å map. Further 3D classification using a smaller angular step enabled us to identify a subset of about 260,000 particle projections that were used to obtain a map with a global indicated resolution of 4.2 Å (Extended Data Figs. 5, 6a–d, Extended Data Table 1). Conformational heterogeneity in this subset was probed using multi-body refinement²⁷, which showed that approximately 40% of the variance in the rotations and translations between NTSR1 and βarr1(ΔCT) is accounted for by two eigenvectors (Extended Data Fig. 6e–h). However, local resolution could not be further improved.

Structure of the NTSR1-βarr1(ΔCT) complex

The 4.2 Å map (Fig. 1a) was sufficient to build a model of the complex (Fig. 1b), with the exception of the N terminus of the receptor (residues 1–49), parts of ICL3 (residues 273–285) and part of the C terminus of the receptor. The transmembrane helices of NTSR1 have well-defined density, and most regions of βarr1(ΔCT) were well-resolved, with the exception of the flexible end of the truncated C terminus (residues 352–382) and parts of the C-edge loops (residues 332–340). The receptor engages βarr1(ΔCT) through a portion of its C terminus, the transmembrane core and the C-terminal end of ICL3. The C terminus of NTSR1 binds the N-lobe groove of βarr1(ΔCT), which contains basic residues that

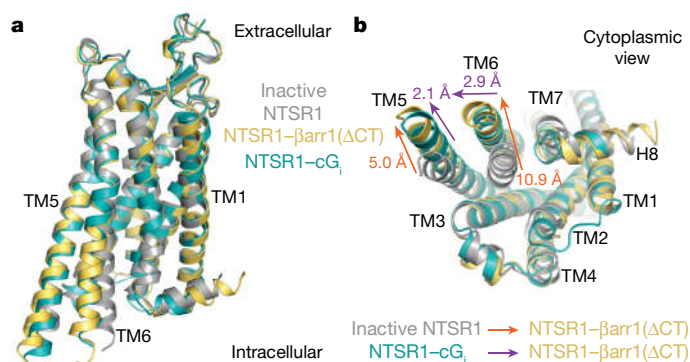


Fig. 2 | β arr1(Δ CT)-bound NTSR1 adopts a conformation similar to that of G_i -bound states. **a, **b**, Side view (**a**) and cytoplasmic view (**b**) showing the overlay of β arr1(Δ CT)-bound NTSR1 (yellow) with inactive NTSR1 (grey), and the canonical G_i -bound NTSR1 structures (cG_i , teal). TM5 in the NTSR1- β arr1(Δ CT) complex has a larger outward displacement than in either NTSR1- G_i complex, while TM6 adopts a conformation similar to that in the canonical NTSR1- G_i complex. Outward motions in TM5 and TM6 of NTSR1, from the inactive to the β arr1(Δ CT)-bound state and from the cG_i -bound to the β arr1(Δ CT)-bound state, are indicated by orange and purple arrows, respectively.**

are known to interact with phosphorylated receptor residues²⁸. The receptor core engages the β arr1(Δ CT) finger loop; this interaction is a critical determinant of arrestin coupling. We also observe a strong density above the β arr1(Δ CT) C-lobe, in the region that was previously shown to be an inositol phosphate binding site^{29–31}. Given that we added dioctyl-phosphatidylinositol-4,5-bisphosphate (diC8-PtdIns(4,5)P₂) during the phosphorylation step, and its presence was confirmed by mass spectroscopy analysis of the intact complex (Supplementary Fig. 1), we conclude that this density corresponds to diC8-PtdIns(4,5)P₂. The functional role of the binding of PtdInsP₂ is discussed below.

We also observe that the C-edge—that is, parts of the 340-loop (residues 330–340) and 191-loop (residues 186–198)—seems to be in contact with the detergent micelle. In Arr1, these loops—named 344-loop and 197-loop—have been shown to interact with the membrane upon arrestin activation³². Notably, and as previously discussed^{32,33}, only Arr1, Arr4 and the long splice variant of β arr1 have the 344- or the 340-loop, whereas β arr2 and the short splice variant of β arr1 do not possess this loop region. No substantial difference in the ability of these isoforms to bind to the β_2 AR and M2 receptors was observed³⁴; however, this loop region may be important in regulating the strength of engagement of

arrestin upon activation, and seems to have an important role in stabilizing our complex. The interaction between the micelle and the C-edge leads to a tilt of β arr1(Δ CT) relative to the receptor (Fig. 1a, b), which may be exaggerated by virtue of the small size of a detergent micelle compared with a planar bilayer. The implications of this flexibility and the role of membrane curvature are discussed below.

Notably, β arr1(Δ CT) engages NTSR1 in a relative orientation that is rotated by approximately 85° in the membrane plane compared with how Arr1 engages rhodopsin (Fig. 1c, Extended Data Fig. 7a). As discussed below, even though both complexes use the same GPCR/arrestin interface, differences in the phosphorylation-mediated interactions, the engagement of the receptor core and the binding of PtdIns(4,5)P₂ could explain this substantial difference. We suggest that this highlights the exceptional plasticity of arrestin and illustrates the spectrum of conceivable receptor–arrestin assemblies. It should be noted that it has previously been anticipated that arrestins may adopt distinct, receptor-specific orientations²⁸.

Conformational changes in NTSR1

The hallmark feature of the activation of family A GPCRs—that is, the outward movement of TM6—is clearly observed in NTSR1 (Fig. 2). Compared with the previously published structure of inactive NTSR1, both TM6 and TM5 in the NTSR1- β arr1(Δ CT) structure move away from the core by 10.9 Å and 5.0 Å, respectively (Fig. 2b), adopting an active conformation similar to that seen in the C-state of the G_i -coupled NTSR1 structure (discussed above)⁹, with a root-mean-square deviation of 0.67 Å. Whereas the displacement of TM6 was similar to that seen in both G_i -bound states, TM5 is moved further outward, by 2.1 Å (by measuring the C α of A270) (Fig. 2b).

Conformational changes in β arr1(Δ CT)

β arr1(Δ CT) bound to NTSR1 adopts a structure that contains the conformational hallmarks of arrestin activation³; we propose this on the basis of a comparison with crystal structures of the inactive β arr1³⁵, the active-state β arr1 bound to a phosphopeptide derived from the vasopressin receptor C-tail and a stabilizing Fab³⁶ (referred to as V2Rpp- β arr1), and the active-state Arr1 bound to rhodopsin¹⁴ (referred to as Rho-Arr1). First, we observe an interdomain twist—that is, a rotation of the C-lobe relative to the N-lobe (Fig. 3a)—of around 16°, which is smaller than the twist of approximately 22° that is observed in the structures of V2Rpp- β arr1 and Rho-Arr1 (Extended Data Fig. 7b). Second, the finger, gate and middle loops—which form the central crest and are essential in receptor coupling—adopt active-state conformations

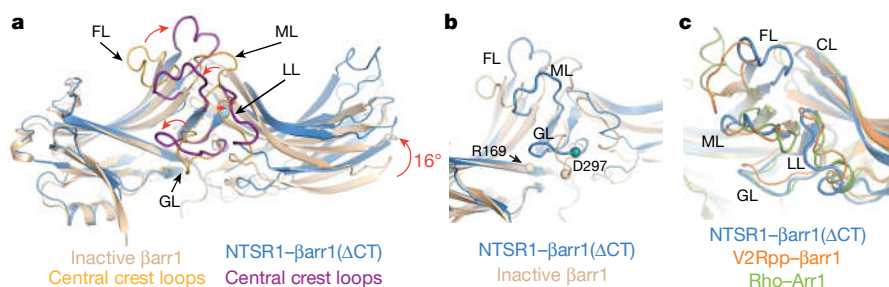


Fig. 3 | NTSR1-bound β arr1(Δ CT) shows activation hallmarks, with some loops in distinct conformations. **a, Overlay of NTSR1-bound β arr1(Δ CT) (blue) with inactive β arr1 (wheat). NTSR1-bound β arr1(Δ CT) undergoes a 16° interdomain twist, a feature observed to varying degrees in all active-state arrestin structures (Supplementary Fig. 8a), and substantial rearrangements of the central crest loops (highlighted in purple and orange, respectively). FL, finger loop; GL, gate loop; LL, lariat loop; ML, middle loop. **b**, The position of**

residue D297 (shown as C α sphere) in the gate loop of the NTSR1- β arr1(Δ CT) complex is too far away from R169 (shown as C α sphere) to stabilize the polar core in the inactive state. **c**, Compared with the active state Rho-bound Arr1 (green) and V2Rpp-bound β arr1 (orange), the middle loop, gate loop and lariat loop of NTSR1-bound β arr1(Δ CT) adopt similar conformations, while the finger loop and the C-loop (CL) adopt more distinct conformations, probably due to the unique receptor–arrestin orientation that we observe.

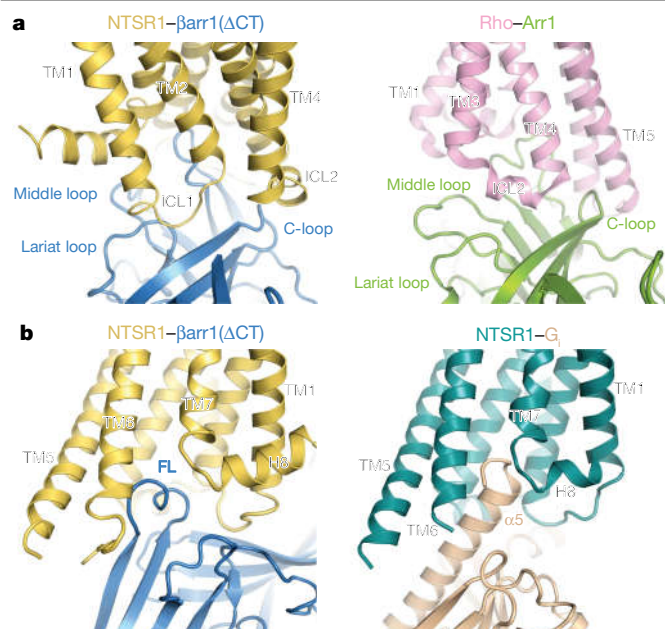


Fig. 4 | Comparison of receptor/arrestin and receptor/G-protein interfaces. **a**, A comparison of the NTSR1/βarr1(ΔCT) interaction interface (left) with the Rho/Arr1 interaction interface (right) highlights the conformational plasticity of arrestin and the 85° rotation between how Arr1 and βarr1(ΔCT) engage rhodopsin and NTSR1, respectively. The same arrestin elements interact with distinct regions on the receptors to form a unique interface. **b**, Engagement of the receptor core by the finger loop of the arrestin (left) and by the Gα,α5-helix (right).

(Fig. 3a). Third, we observe that the polar core is disrupted, as evidenced by the backbone position of residue D297 in the gate loop being moved away from R169, abrogating a key contact that stabilizes the inactive state (Fig. 3b). Moreover, because βarr1(ΔCT) is truncated at residue 382, residue R393 is no longer present to contribute to the formation of the polar core. Notably, the backbones of the gate loop and middle loop in our structure overlay well with those of the Rho–Arr1 and V2Rpp–βarr1 complexes, whereas the finger loop and the C-loop adopt distinct conformations (Fig. 3c).

NTSR1/βarr1(ΔCT) interaction interface

Several interaction elements stabilize the NTSR1–βarr1(ΔCT) complex, and they probably explain the unique orientation that we observe compared with the Rho–Arr1 complex (Fig. 1c). Although the finger loop, C-loop, lariat loop and middle loop of Arr1 and βarr1(ΔCT) are part of the interface, they interact with different regions of rhodopsin and NTSR1, respectively, resulting in very different orientations of arrestin in the two complexes (Fig. 4a). In the NTSR1–βarr1(ΔCT) complex, the finger loop adopts a helical structure—which is also seen for Arr1³⁷—and it inserts into the receptor intracellular cavity, but it is extended farther away from the N-lobe of arrestin relative to the finger loop of Arr1 in the rhodopsin complex¹⁴ (Figs. 3c, 4a). The finger loop of βarr1(ΔCT) occupies a position similar to that of the α5-helix of Gα_i in the NTSR1–G_i complex (Fig. 4b). In the NTSR1–βarr1(ΔCT) complex, ICL2 forms a helix and is located on the outer side of the C-loop, whereas in the Rho–Arr1 complex ICL2 is sandwiched between the C-loop, middle loop and lariat loop (Figs. 4a, 5a). It is notable that the ICL2 of rhodopsin does not form a helix in the G_i-bound structure, whereas it does in the Arr1-bound structure; this is in contrast to NTSR1, in which ICL2 adopts a very similar conformation in its G_i-bound and βarr1(ΔCT)-bound states.

NTSR1 phosphorylation stabilizes the complex

As noted above, receptor phosphorylation is essential for the formation of a stable complex (Extended Data Fig. 2). Phosphoproteomics experiments on a pre-formed NTSR1–βarr1(ΔCT) complex identified four phosphorylation sites in ICL3 and six in the C terminus (Extended Data Fig. 8a, Supplementary Figs. 2–5, Supplementary Tables 8–39), with less than full occupancy of each position. We propose that the heterogeneity we observe is the result of distributive phosphorylation by GRK5 *in vitro*—consistent with literature reports stating that GRKs have limited sequence recognition constraints²³—and might even be important to enable arrestin to adopt a range of conformational states.

We observed density corresponding to part of the C terminus of NTSR1 in the N-lobe of βarr1(ΔCT) (Fig. 5b), with a distinct bulge in the density adjacent to K294 of βarr1(ΔCT) consistent with a phosphate; however, owing to the resolution we cannot definitively determine the register of the C terminus of NTSR1 within this density. On the basis of the distance between the last amino acid observed for H8 and the density observed for the C-terminal peptide, we can probably exclude S396, S401, S403 and S404 from being at this position. As such, it is possible that our structure represents pT407 making a contact to K294, and we have modelled it as such in Fig. 5b. Both the V2Rpp–βarr1 and the Rho–Arr1 complexes have a phosphate in the same position, interacting with the gate loop lysine and thereby stabilizing arrestin in the active conformation^{14,36} (Extended Data Fig. 8b).

We also observe strong density for the C-terminal end of ICL3, which turns sharply at the end of TM6 and extends behind TM5 to the base of the finger loop (Fig. 5c, d). The density for residue S287 of ICL3, which is phosphorylated in our complex, is adjacent to residues R76 and K77 of βarr1(ΔCT) at the base of the finger loop. We speculate that phosphorylation of S287, but potentially any phosphorylated residue in the inherently flexible ICL3, could serve to disrupt the contact between K77 and E313 that maintains arrestin in the inactive conformation³⁸. It is noteworthy that, in the V2Rpp–βarr1 structure, K77 is stabilized in the active state by pT347 of the V2Rpp³⁶ (Extended Data Fig. 8b).

PtdIns(4,5)P₂ stabilizes NTSR1–βarr1(ΔCT)

In the NTSR1–βarr1(ΔCT) structure, arrestin is strongly tilted towards the membrane, at an angle of about 40° relative to the membrane plane (Fig. 1a, b); this same angle is only about 15° for Rho–Arr1. The difference in tilt may be attributed to the interaction of the C-edge with the detergent micelle, as well as to the presence of PtdIns(4,5)P₂ bridging the membrane surface of TM1 and TM4 with the top of the C-lobe of arrestin. Although the observation of a PtdIns(4,5)P₂ molecule bridging NTSR1 and βarr1(ΔCT) was unexpected, this region of βarr1 and βarr2 was previously shown to bind inositol phosphates^{29–31}, and native mass spectrometry experiments had shown that NTSR1 was capable of binding to PtdIns(4,5)P₂ through contacts between positively charged residues on the membrane-facing side of TM1 and TM4³⁹. We confirmed the presence of PtdIns(4,5)P₂ in our complex by mass spectrometry, and determined that a fluorescent PtdIns(4,5)P₂ analogue could bind to both NTSR1 (dissociation constant, $K_d = 0.3 \mu\text{M}$) and βarr1(ΔCT) ($K_d = 0.9 \mu\text{M}$) (Supplementary Fig. 1). PtdIns(4,5)P₂ fits well into the density observed between the detergent micelle and the C-lobe of βarr1(ΔCT) (Fig. 5e, f). Although the precise orientation of the phosphatidylinositol head group is ambiguous at the resolution of our map, the modelled position places the phosphates at positions 4 and 5 within range to interact with R236, K250, K324 and K326 of βarr1(ΔCT). The hydroxyl group at position 3 of the PtdIns(4,5)P₂ head group can form a hydrogen bond with R182 on TM4 of the receptor, and the bridging phosphate can form a hydrogen bond with Y103 on TM2 (Fig. 5f). Using a NanoBiT assay⁴⁰, we examined in cells the effect of mutating residues that are involved in the binding of PtdIns(4,5)P₂ to arrestin^{29,38}. When stimulated by an agonist, the mutant βarr1(K232Q/

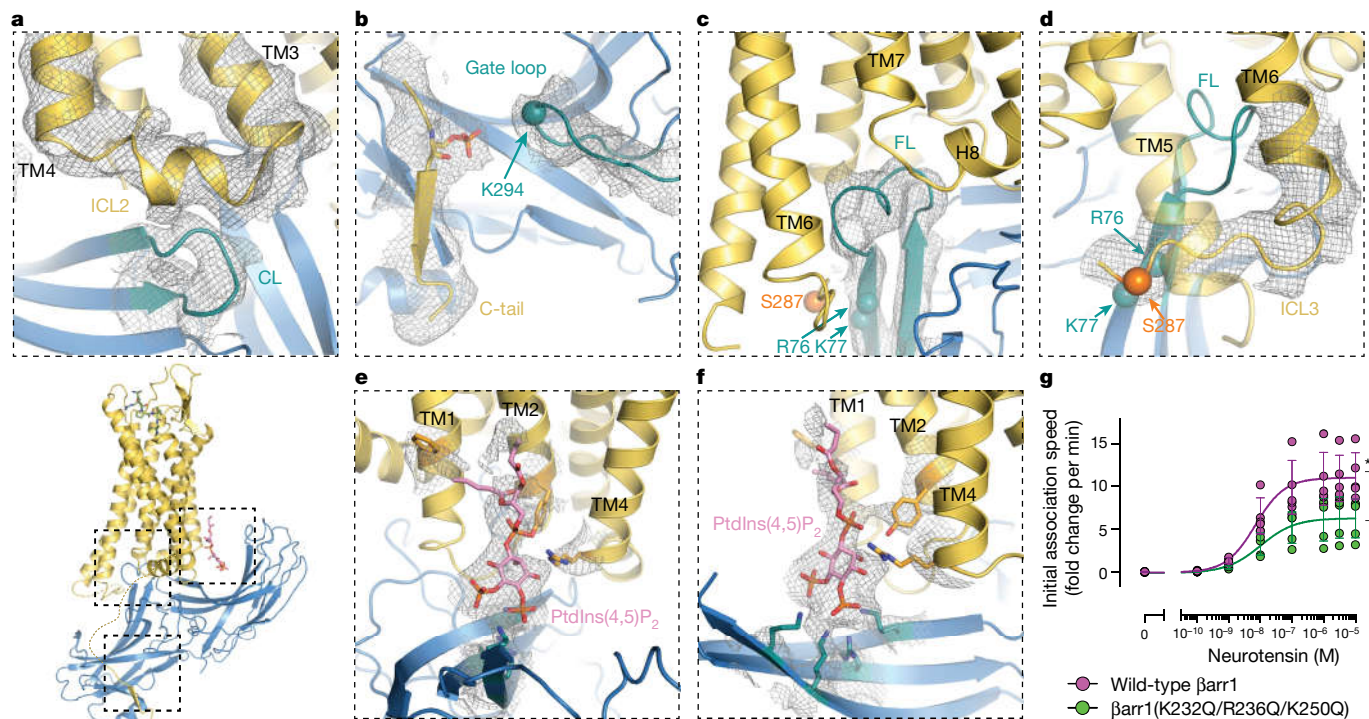


Fig. 5 | The NTSR1/βarr1(ΔCT) interface comprises several distinct types of interactions. Multiple interaction interfaces (dashed boxes) stabilize the NTSR1–βarr1(ΔCT) complex. **a–f**, Close-up view of distinct interaction interfaces, with local density from the electron microscopy map shown as grey mesh. **a**, The C-loop (teal) is in close proximity to ICL2, possibly interacting through hydrophobic contacts. **b**, The gate loop is in close proximity to part of the C-tail of NTSR1. A bulge in the map density suggests that a phosphorylated residue would be positioned as shown to interact with K294, thus disrupting the polar core. **c**, The finger loop (teal) inserts into the receptor cavity formed by outward motion of TM6. **d**, The C-terminal part of ICL3 adopts a distinct orientation, forming a sharp angle with the intracellular end of TM6. A charge–charge interaction between a phosphorylated residue, such as S287,

and positively charged residues at the base of the finger loop (R76 and K77) may explain this unusual loop conformation. **e, f**, A PtdIns(4,5)P₂ molecule interacts with basic residues of the βarr1(ΔCT) C-lobe groove and residues on the membrane-facing side of TM1, TM2 and TM4. **g**, A NanoBIT-based proximity assay shows that the triple mutant βarr1(K232Q/R236Q/K250Q), in which residues required for the binding of PtdIns(4,5)P₂ are mutated, is recruited to activated NTSR1 to a lesser extent (**P* = 0.02) than wild-type βarr1. Data points and error bars represent data from individual experiments and s.d., respectively. The maximum response (*E*_{max}) of NTS-stimulated βarr1 recruitment to NTSR1 is 11.2 ± 2.8 for wild-type βarr1 and 6.4 ± 2.4 for βarr1(K232Q/R236Q/K250Q). One-way analysis of variance, *n* = 5 biologically independent experiments.

R236Q/K250Q)—containing lysine-to-glutamine substitutions at residues 232 and 250, and an arginine-to-glutamine substitution at residue 236—showed an approximately 40% reduction in recruitment of arrestin to NTSR1 when compared with the wild-type (Fig. 5g). This is consistent with the binding of arrestin to PtdIns(4,5)P₂ being important for complex formation with NTSR1.

Diversity of GPCR–arrestin interactions

We present the structure of the NTSR1–βarr1(ΔCT) complex using a native receptor that is phosphorylated *in vitro* by GRK5. A comparison with the structures of the Rho–Arr1 complex and with the M2R–βarr1 complex in lipid nanodiscs—described in a related study⁴¹—highlights the plasticity and diversity of interactions that comprise the receptor/arrestin interface.

We found that the binding of PtdIns(4,5)P₂ at the NTSR1/βarr1(ΔCT) interface stabilizes the complex and may impose a stronger arrestin tilt relative to the membrane plane compared to that in M2R–βarr1 and Rho–Arr1. This tilt may be exaggerated by the smaller hydrophobic interface that a detergent micelle provides compared with a planar lipid bilayer. Conformational analysis of the cryo-EM projections indicates the least-inclined state captured is tilted to a comparable extent to that seen in Rho–Arr1 and M2R–βarr1. However, even the most-tilted orientation may be meaningful in the context of non-planar membrane structures that can feature high degrees of curvature, such as endosomes, from which GPCR–arrestin signalling can also

occur⁴² (Extended Data Fig. 9). The observations that interactions with membrane phosphoinositides enable arrestins to maintain signalling-competent conformations after dissociating from GPCRs (referred to as ‘action at a distance’)—and that these dissociated arrestins, as well as GPCR–arrestin complexes, can signal from clathrin-coated structures^{38,43}—further supports the notion that PtdIns(4,5)P₂ and membrane curvature may have a role in arrestin function. As such, once arrestin is activated by the receptor core, C-tail and/or ICL3 engagement, the binding of specific lipids such as PtdIns(4,5)P₂, and the insertion of the C-edge into the lipid bilayer may contribute to the maintenance of an active conformation that no longer requires the presence of the receptor. It has been shown that NTSR1–G_q mediated signalling activates phospholipase C, and the production of inositol phosphate and diacylglycerol from PtdIns(4,5)P₂, as detailed in a previous publication¹⁸. Thus, NTSR1 signalling through G_q can change the membrane content of PtdIns(4,5)P₂ and possibly influence interactions between NTSR1 and βarr1.

Compared with both Rho–Arr1 and M2R–βarr1, the relative orientation between the receptor and arrestin that we observe for NTSR1–βarr1(ΔCT) is unique. This suggests that the same arrestin regions can engage receptors in different ways, and that the orientation is modulated by the nature of the interactions present. For example, the interaction with the phosphorylated ICL3 as well as binding to PtdIns(4,5)P₂—which are observed only in the NTSR1–βarr1(ΔCT) complex—can be satisfied only by adopting a different relative orientation of arrestin. Within arrestin we observe different conformations of the finger loop

between NTSR1- β arr1(Δ CT), M2R- β arr1 and Rho-Arr1, suggesting that this loop can be ordered in different ways; this may be important for the recognition of different receptors. The ability of arrestin to assume multiple conformations^{44,45}, including different twist angles and relative orientations, might be important in enabling binding to differently phosphorylated receptors. Taken together, this high degree of conformational plasticity and the multiplicity of interface contacts might be what enables only two β -arrestin isoforms to interact with a large number of GPCRs that share little sequence homology at the arrestin/receptor interface.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1953-1>.

- Scheerer, P. & Sommer, M. E. Structural mechanism of arrestin activation. *Curr. Opin. Struct. Biol.* **45**, 160–169 (2017).
- Hilger, D., Masureel, M. & Kobilka, B. K. Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).
- Rankovic, Z., Brust, T. F. & Bohn, L. M. Biased agonism: an emerging paradigm in GPCR drug discovery. *Bioorg. Med. Chem. Lett.* **26**, 241–250 (2016).
- Luttrell, L. M., Maudsley, S. & Bohn, L. M. Fulfilling the promise of “biased” G protein-coupled receptor agonism. *Mol. Pharmacol.* **88**, 579–588 (2015).
- Barak, L. S. et al. ML314: a biased neurotensin receptor ligand for methamphetamine abuse. *ACS Chem. Biol.* **11**, 1880–1890 (2016).
- Peterson, Y. K. & Luttrell, L. M. The diverse roles of arrestin scaffolds in G protein-coupled receptor signaling. *Pharmacol. Rev.* **69**, 256–297 (2017).
- Laporte, S. A. & Scott, M. G. H. β -Arrestins: multitask scaffolds orchestrating the where and when in cell signalling. *Methods Mol. Biol.* **1957**, 9–55 (2019).
- Krishna Kumar, K. et al. Structure of a signaling cannabinoid receptor 1-G protein complex. *Cell* **176**, 448–458.e12 (2019).
- Kato, H. E. et al. Conformational transitions of a neurotensin receptor 1-G_i complex. *Nature* **572**, 80–85 (2019).
- Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G. & Kobilka, B. K. Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science* **364**, 552–557 (2019).
- Gao, Y. et al. Structures of the rhodopsin-transducin complex: insights into G-protein activation. *Mol. Cell* **75**, 781–790 (2019).
- García-Nafria, J. & Tate, C. G. Cryo-EM structures of GPCRs coupled to G_s, G_i and G_o. *Mol. Cell. Endocrinol.* **488**, 1–13 (2019).
- Glukhova, A. et al. Rules of engagement: GPCRs and G proteins. *ACS Pharmacol. Transl. Sci.* **1**, 73–83 (2018).
- Zhou, X. E. et al. Identification of phosphorylation codes for arrestin recruitment by G protein-coupled receptors. *Cell* **170**, 457–469.e13 (2017).
- Yang, F. et al. Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and ¹⁹F-NMR. *Nat. Commun.* **6**, 8202 (2015).
- Lohse, M. J., Benovic, J. L., Codina, J., Caron, M. G. & Lefkowitz, R. J. β -Arrestin: a protein that regulates β -adrenergic receptor function. *Science* **248**, 1547–1550 (1990).
- Rostène, W. H. & Alexander, M. J. Neurotensin and neuroendocrine regulation. *Front. Neuroendocrinol.* **18**, 115–173 (1997).
- Besserer-Offroy, É. et al. The signaling signature of the neurotensin type 1 receptor with endogenous ligands. *Eur. J. Pharmacol.* **805**, 1–13 (2017).
- White, J. F. et al. Structure of the agonist-bound neurotensin receptor. *Nature* **490**, 508–513 (2012).
- Egloff, P. et al. Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **111**, E655–E662 (2014).
- Krumm, B. E., White, J. F., Shah, P. & Grishammer, R. Structural prerequisites for G-protein activation by the neurotensin receptor. *Nat. Commun.* **6**, 7895 (2015).
- Komolov, K. E. et al. Structural and functional analysis of a β_2 -adrenergic receptor complex with GRK5. *Cell* **169**, 407–421.e16 (2017).
- Komolov, K. E. & Benovic, J. L. G protein-coupled receptor kinases: Past, present and future. *Cell. Signal.* **41**, 17–24 (2018).
- Inagaki, S. et al. G protein-coupled receptor kinase 2 (GRK2) and 5 (GRK5) exhibit selective phosphorylation of the neurotensin receptor in vitro. *Biochemistry* **54**, 4320–4329 (2015).
- Vishnivetskiy, S. A. et al. An additional phosphate-binding element in arrestin molecule. Implications for the mechanism of arrestin activation. *J. Biol. Chem.* **275**, 41049–41057 (2000).
- Peisley, A. & Skiniotis, G. 2D projection analysis of GPCR complexes by negative stain electron microscopy. *Methods Mol. Biol.* **1335**, 29–38 (2015).
- Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
- Gurevich, V. V. & Gurevich, E. V. The structural basis of arrestin-mediated regulation of G-protein-coupled receptors. *Pharmacol. Ther.* **110**, 465–502 (2006).
- Gaidarov, I., Krupnick, J. G., Falck, J. R., Benovic, J. L. & Keen, J. H. Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. *EMBO J.* **18**, 871–881 (1999).
- Chen, Q. et al. Structural basis of arrestin-3 activation and signaling. *Nat. Commun.* **8**, 1427 (2017).
- Milano, S. K., Kim, Y. M., Stefano, F. P., Benovic, J. L. & Brenner, C. Nonvisual arrestin oligomerization and cellular localization are regulated by inositol hexakisphosphate binding. *J. Biol. Chem.* **281**, 9812–9823 (2006).
- Lally, C. C., Bauer, B., Selent, J. & Sommer, M. E. C-edge loops of arrestin function as a membrane anchor. *Nat. Commun.* **8**, 14258 (2017).
- Kang, D. S. et al. Structure of an arrestin2-clathrin complex reveals a novel clathrin binding domain that modulates receptor trafficking. *J. Biol. Chem.* **284**, 29860–29872 (2009).
- Gurevich, V. V. et al. Arrestin interactions with G protein-coupled receptors. Direct binding studies of wild type and mutant arrestins with rhodopsin, β_2 -adrenergic, and m2 muscarinic cholinergic receptors. *J. Biol. Chem.* **270**, 720–731 (1995).
- Han, M., Gurevich, V. V., Vishnivetskiy, S. A., Sigler, P. B. & Schubert, C. Crystal structure of β -arrestin at 1.9 Å: possible mechanism of receptor binding and membrane translocation. *Structure* **9**, 869–880 (2001).
- Shukla, A. K. et al. Structure of active β -arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature* **497**, 137–141 (2013).
- Szcepek, M. et al. Crystal structure of a common GPCR-binding interface for G protein and arrestin. *Nat. Commun.* **5**, 4801 (2014).
- Eichel, K. et al. Catalytic activation of β -arrestin by GPCRs. *Nature* **557**, 381–386 (2018).
- Yen, H. Y. et al. PtdIns(4,5)P₂ stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **559**, 423–427 (2018).
- Inoue, A. et al. Illuminating G-protein-coupling selectivity of GPCRs. *Cell* **177**, 1933–1947.e25 (2019).
- Staus, D. P. et al. Structure of the M2 muscarinic receptor- β -arrestin complex in a lipid nanodisc. *Nature* <https://www.doi.org/10.1038/s41586-020-1954-0> (2020).
- Eichel, K. & von Zastrow, M. Subcellular organization of GPCR signaling. *Trends Pharmacol. Sci.* **39**, 200–208 (2018).
- Eichel, K., Jullié, D. & von Zastrow, M. β -Arrestin drives MAP kinase signalling from clathrin-coated structures after GPCR dissociation. *Nat. Cell Biol.* **18**, 303–310 (2016).
- Lee, M. H. et al. The conformational signature of β -arrestin2 predicts its trafficking and signalling functions. *Nature* **531**, 665–668 (2016).
- Nuber, S. et al. β -Arrestin biosensors reveal a rapid, receptor-dependent activation/deactivation cycle. *Nature* **531**, 661–664 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data analysis and figure preparation

Figures were created using the PyMOL 2.3.2 Molecular Graphics System (Schrödinger), and the UCSF Chimera X 0.9 package⁴⁶. Data were plotted with GraphPad Prism 8.3.0. Mass spectrometry data were analysed with Byonic v.3.4 and Byologic v.3.4-55 (Protein Metrics), DataAnalysis v.5.2 (Bruker Daltonics) and Excalibur v.4.1 (Thermo Fisher Scientific).

Screening GPCRs for strong arrestin coupling

GPCRs with an N-terminal Flag tag and a C-terminal eGFP followed by deca-His and RhoID4 tags were constructed to pCDNA3.1(+) and expressed in HEK293T cells (ATCC, confirmed mycoplasma negative by manufacturer). Cells were transfected with Lipofectamine 2000 (Invitrogen) and incubated for 72 h before being collected. Cell pellets were solubilized in solubilization buffer (1% lauryl maltose neopentyl glycol (LMNG, Anatrace), 0.1% cholesteryl hemisuccinate tris salt (CHS, Steraloids), 500 mM NaCl, 20 mM HEPES 7.4, 25% glycerol, protease inhibitors and 1 μ M ligand) for 2 h at 4 °C. After centrifugation, the supernatant was incubated with Ni-NTA for 30 min at 4 °C. The resin was washed in batch with 5 column volumes of wash buffer (0.01% LMNG, 0.001% CHS, 20 mM HEPES pH 7.4, 500 mM NaCl, 10 mM imidazole and 1 μ M ligand). The receptor was eluted in wash buffer supplemented with 200 mM imidazole. To test arrestin coupling, rat β arr1(1–392) was activated by incubating with V2Rpp and Fab30 before the addition of receptor. The V2Rpp peptide: ARGRpTPpSLGPQDEpSCpTpTapSpSLAKDTSS was obtained by custom peptide synthesis (Tufts University Core Facility). Fab30 was expressed and purified as previously described³⁶. Pre-activated β arr1 (5 μ M) was then added to each receptor and incubated for 30 min at room temperature. Samples were then subjected to size-exclusion chromatography (SEC) using a Superdex 200 10/300 GL column (GE Healthcare) and monitored by the fluorescence of eGFP (excitation wavelength (λ_{ex}) 480 nm, emission wavelength (λ_{em}) 512 nm).

NTSR1 expression and purification

Full-length wild-type human NTSR1 was modified with an N-terminal Flag tag followed by an octa-histidine tag, and cloned into pFastBac1 vector. NTSR1 was expressed in *Spodoptera frugiperda* (Sf9) insect cells (Expression Systems) using a FastBac-derived baculovirus. Cells were infected at a density of 4×10^6 cells per ml and collected 60 h post infection. Cells were lysed in hypotonic buffer (10 mM HEPES, pH 7.4, and protease inhibitors) and solubilized at 4 °C for 2 h in a buffer containing 1% LMNG, 0.1% CHS, 0.3% sodium cholate (Sigma), 20 mM HEPES 7.4, 500 mM NaCl, 25% glycerol, iodoacetamide (to cap cysteine residues) and protease inhibitors. Insoluble debris was removed by centrifugation and the supernatant was incubated with Ni-NTA (Qia-gen) resin for 1 h at 4 °C. The resin was washed in batch with buffer containing 0.01% LMNG, 0.001% CHS, 0.003% sodium cholate, 20 mM HEPES pH 7.4, 500 mM NaCl, 10 mM imidazole and eluted with the same buffer supplemented with 200 mM imidazole, 2 mM CaCl_2 and 10 μ M NTS_{8–13} (acetate salt, Sigma). The eluate was loaded onto M1–Flag immunoaffinity resin and washed with buffer containing 0.01% LMNG, 0.001% CHS, 0.003% sodium cholate, 20 mM HEPES pH 7.4, 500 mM NaCl, 10 mM imidazole, 0.1 μ M NTS_{8–13} and 2 mM CaCl_2 . The receptor was eluted with buffer containing 100 mM NaCl, 20 mM HEPES pH 7.4, 0.005% LMNG, 0.005% CHS, 1 μ M NTS_{8–13}, 0.2 mg ml^{–1} Flag peptide (DYKDDDDK) and 5 mM EDTA. Elution fractions containing receptor were pooled and subjected to polishing by size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) in 20 mM HEPES, pH 7.4, 100 mM NaCl, 0.0025% LMNG, 0.00025% CHS and 0.1 μ M NTS_{8–13}. Peak fractions were pooled and concentrated to 200 μ M and aliquots were flash-frozen and stored at –80 °C until use.

GRK5 expression and purification

Full-length wild-type human GRK5 was modified with a C-terminal hexa-histidine tag and cloned into pVL1392 vector for baculovirus production. GRK5 was expressed and purified as previously described⁴⁷. In brief, Sf9 insect cells (Expression Systems) were infected with a BestBac-derived baculovirus at a density of 3.5×10^6 cells per ml and collected 48 h post-infection. Cells were resuspended, lysed by sonication and the supernatant was applied to Ni-NTA resin. The resin was washed with lysis buffer and GRK5 was eluted with lysis buffer supplemented with 200 mM imidazole. The combined eluate was then subjected to cation-exchange chromatography using a MonoS10/100 column (GE Healthcare) and eluted with a linear gradient of NaCl. Fractions containing GRK5 were combined and run on a Superdex 200 10/300 GL column (GE Healthcare). GRK5 was aliquoted, flash-frozen, and stored at –80 °C until use.

Arrestin expression and purification

The long splice variant of human, cysteine-free (C59V/C125S/C140L/C150V/C242V/C251V/C269S) β -arrestin 1 was used as full-length arrestin for these studies, and modified with an N-terminal hexa-histidine tag, followed by a 3C protease site, a GG linker, AviTag and GSGGS linker. The sequence was codon-optimized for expression in *Escherichia coli* and cloned into a pET-15b vector. β arr1(Δ CT) was prepared by using a β arr1 construct truncated at residue 382. All arrestin constructs used were prepared as follows: NiCo21(DE3) competent *E. coli* (New England Biolabs) were transformed and large-scale cultures were grown in TB + ampicillin at 37 °C until an optical density at 600 nm (OD_{600}) of 1.0 was reached. Cells were then transferred to room temperature and induced with 25 μ M IPTG when the OD_{600} reached 2.0. Cells were collected 20 h post-induction and resuspended in lysis buffer (50 mM HEPES pH 7.4, 500 mM NaCl, 15% glycerol, 7.13 mM β -mercaptoethanol) to a final volume of 40 ml per litre of cells. Cells were lysed by sonication and the clarified lysate applied to nickel sepharose and batch incubated for 1.5 h at 4 °C. The resin was washed with 10 column volumes of wash buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 10% glycerol, 7.13 mM β -mercaptoethanol) + 20 mM imidazole, followed by 10 column volumes of wash buffer + 40 mM imidazole. The protein was then eluted with 5 column volumes of wash buffer + 200 mM imidazole and dialysed overnight in 100 \times volume of dialysis buffer (20 mM HEPES 7.4, 200 mM NaCl, 2 mM β -mercaptoethanol, 10% glycerol) in the presence of 1:10 (w/w) of 3C protease. The digested protein was then subjected to reverse-nickel purification and diluted with dialysis buffer containing no NaCl to bring the NaCl concentration to 75 mM. The protein was then purified by ion-exchange chromatography (mono Q 10/100 GL, GE Healthcare), followed by size-exclusion using a Superdex 200 Increase 10/300 GL column (GE Healthcare) with SEC buffer (20 mM HEPES pH 7.4, 300 mM NaCl, 10% glycerol). The purified protein was then concentrated to 200–300 μ M using a 30-kDa spin concentrator and aliquots were flash-frozen in liquid nitrogen and stored at –80 °C until use.

NTSR1 phosphorylation

NTSR1 (2.5 μ M) was equilibrated in phosphorylation buffer (20 mM bis-tris propane (BTP) pH 7.5, 35 mM NaCl, 5 mM MgCl_2 , 20 μ M NTS_{8–13}, 20 μ M diC8-PtdIns(4,5)P₂, 0.05 mM TCEP, 0.002% MNG, 0.0002% CHS) at 25 °C with gentle mixing for 1 h. GRK5 was added to the reaction to a final concentration of 200 nM, and briefly incubated while the reaction was warmed from 25 °C to 30 °C. ATP was added to a final concentration of 1 mM. The reaction was sampled at various time points by removing a small volume, mixing it with EDTA (to 50 mM final) to halt phosphorylation and evaluated by ion-exchange chromatography, which was performed as follows: 10 μ l of reaction mixture was added to 40 μ l of buffer A (20 mM BTP pH 8.6, 50 mM NaCl, 0.004% LMNG) and injected onto an equilibrated (100% buffer A) monoQ (5/50) cation exchange

column (GE Healthcare). The column was washed with 1 column volume (CV) of a linear gradient from 0 to 10% buffer B (20 mM BTP pH 8.6, 1 M NaCl, 0.004% LMNG), 6 CV at 10% B, 1 CV from 10 to 40% B then 6 CV at 40% B before being re-equilibrated into buffer A. In some cases, a longer linear gradient was used instead of a stepped elution. This involved 1 CV from 0% B to 5% B, 5 CV 5% B then 20 CV from 5 to 40% B. Protein elution was monitored by tracking tryptophan fluorescence at λ_{ex} of 280 nm and λ_{em} of 340 nm. Upon completion, the reaction was supplemented with CaCl_2 to a final concentration of 2 mM and applied to an equilibrated M1–Flag immunoaffinity resin and washed with buffer containing 0.004% LMNG, 0.004% CHS, 20 mM HEPES pH 7.4, 100 mM NaCl, 0.2 μM NTS_{8–13}, 2 mM CaCl_2 . The receptor was eluted with buffer containing 100 mM NaCl, 20 mM HEPES pH 7.4, 0.004% LMNG, 0.004% CHS, 0.2 μM NTS_{8–13}, 0.2 mg ml^{−1} 1× Flag peptide (DYKDDDDK), 5 mM EDTA. Elution fractions containing receptor were pooled and used for arrestin complexing.

Analytical fluorescence size-exclusion chromatography for construct screening

In a final volume of 20 μl , NTSR1 (4.5 μM), the respective arrestin construct (9 μM) and NTS_{8–13} peptide (50 μM) were incubated in buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 0.004% LMNG, 0.0004% CHS and 0.2 μM NTS_{8–13}. Using a Prominence-i LC autosampler (Shimadzu), 10 μl was injected onto a ENrich size-exclusion chromatography 650/10 × 300 column (Bio-Rad) pre-equilibrated in 20 mM HEPES pH 7.4 100 mM NaCl, 0.004% LMNG, 0.004% CHS and 0.2 μM NTS_{8–13}, and run at a flow rate of 0.8 ml min^{−1}. Tryptophan fluorescence was monitored at λ_{ex} of 280 nm and λ_{em} of 340 nm.

Formation and purification of NTSR1– β arr1(ΔCT) complex

For cryo-EM, phosphorylated NTSR1 was mixed in an equimolar ratio with β arr1(ΔCT) and diC8-PtdIns(4,5)P₂ at a concentration of around 5 μM and supplemented with NTS_{8–13} to 10 μM final concentration. The mixture was incubated at 25 °C for 30 min before being concentrated with a 50 kDa molecular weight cut-off concentrator (Amicon or Vivaspin) to around 350 μl and purified by size-exclusion chromatography using two Superdex 200 Increase 10/300 GL columns (GE Healthcare) connected in tandem. The mobile phase used was 20 mM HEPES pH 7.4, 100 mM NaCl, 0.00075% LMNG, 0.00025% GDN 0.0001% CHS and 0.2 μM NTS_{8–13}. Fractions containing complex were combined and diluted to 1 μM final concentration, and sulfo-LC-SDA (also known as sulfo-NHS-LC-diazirine; sulfosuccinimidyl 6-(4,4'-azipentanamido) hexanoate) (Thermo Fisher) as a solution in DMSO was added to 250 μM final concentration, and such that the final DMSO concentration was below 4%. The reaction was allowed to proceed for 45 min at 25 °C in the dark, before hydroxylamine was added to a final concentration of 3 mM and incubated for an additional 15 min. The sample was distributed into a clear 96-well plate (90 μl per well), put on ice and irradiated for 45 min using a UVL-56 lamp. The sample was then pooled and concentrated to around 500 μl then re-run on size-exclusion chromatography, again using two Superdex 200 Increase 10/300 GL columns (GE Healthcare) connected in tandem. Peak fractions were combined and concentrated with a 100 kDa molecular weight cut-off concentrator (Amicon) to a final concentration of 4.5 mg ml^{−1}.

Cryo-EM sample preparation and image acquisition

An aliquot of 3.5 μl NTSR1– β arr1(ΔCT) sample was deposited onto glow-discharged 200 mesh grids (Quantifoil R1.2/1.3) and plunge-frozen into liquid ethane using an FEI Vitrobot Mark IV (Thermo Fisher Scientific). Several sessions of data collection were conducted on the same Titan Krios equipped with an energy filter and operated at 300 keV using a nominal magnification of 130,000×. Movies were captured using a Gatan K2 Summit direct electron detector in counted mode, which resulted in a pixel size of 1.06 Å. Movie stacks were obtained with a defocus range of −1.0 to −2.0 μm , using SerialEM 3.7.10⁴⁸ with a set of

customized scripts enabling automated low-dose image acquisition. Each movie stack was recorded for a total of 8 s with 0.2 s per frame. The exposure rate was seven electrons per pixel per second. Initial datasets were collected by one exposure per hole per single stage movement but subsequent collections used nine-hole exposures per single stage movement with beam-tilt compensated by using multi-record strategy implemented in SerialEM 3.7.10.

Cryo-EM data processing

A total of 18,797 image stacks were subjected to beam-induced motion correction using MotionCor2⁴⁹. Contrast transfer function parameters for each micrograph were estimated from the exposure-weighted averages of all frames by Gctf v.1.06⁵⁰. The following processes were all performed using RELION 3.0⁵¹, except those mentioned specifically. After 2D classification, 2,628,700 particles were divided into 6 subsets for 3D classification. A reference map for 3D classification was generated by the '3D initial model' script in RELION 3.0 using default stochastic gradient descent parameters. All stable classes were then combined for 3D refinement, which led to a 4.5 Å map. Further 3D classification of this particle set resulted in two best classes with a total of 263,965 particles. The resulting Bayesian polished particles were used to build the final map with an overall resolution of 4.2 Å, as determined by Fourier shell correlation (FSC) using a cutoff of 0.143. To probe the conformational dynamics, multi-body refinement was performed on the 4.2 Å map with two bodies corresponding to NTSR1 and β arr1. Local resolution was estimated with the Bsoft 2.0.6 package⁵² (Extended Data Fig. 6b).

Model building and refinement

The initial template for NTSR1– β arr1(ΔCT) was built from the receptor coordinates of the crystal structure of NTSR1–NTS_{8–13} (RCSB Protein Data Bank (PDB): 4GRV) and arrestin coordinates from the V2Rpp– β arr1 structure (PDB: 4JQI). The initial coordinates for the NTS_{8–13} ligand were taken from the NTSR1–NTS_{8–13} crystal structure (PDB: 4GRV). Models were docked into the electron microscopy density map using UCSF Chimera 1.14⁵³, then refined by several iterations of automated refinement in PHENIX interspersed with manual adjustments in Coot 0.8.9⁵⁴. The NTSR1 C terminus was built on the basis of the structure of V2Rpp bound to β arr1 (PDB: 4JQI). The final model was subjected to global refinement and minimization in real space using phenix.real_space_refine in PHENIX v.1.16⁵⁵. MolProbity 4.5 was used to evaluate model geometry⁵⁶ (Extended Data Table 1). FSC curves were calculated between the resulting model and the half map used for refinement as well as between the resulting model and the other half map for cross-validation (Extended Data Fig. 6a).

Phosphoproteomics experiments

Receptor samples were denatured, reduced, alkylated and digested according to the manufacturers' protocols (Protifi). In brief, 5–10 μg of receptor sample in 0.004% LMNG/0.0004% CHS buffer was diluted twofold with buffer containing approximately 10% SDS and 100 mM triethylammonium bicarbonate (TEAB, Sigma) pH 7.6. Cysteine residues were reduced by the addition of freshly dissolved dithiothreitol (Sigma) to 20 mM final concentration and heated to 95 °C for 10 min. Samples were cooled to room temperature and iodoacetamide (Sigma) was added to 40 mM final concentration. Samples were incubated for 30 min with light excluded at 25 °C with gentle agitation using a Thermomixer. Phosphoric acid (12%) was added to lower the pH for binding to the S-trap column, followed by binding buffer: 90% methanol, 100 mM TEAB pH 7.1. The sample was then added to the S-trap and sequentially spun at 4,000g until all material was loaded discarding the flow-through. The column was washed three times with binding buffer, as above. Freshly reconstituted Trypsin/LysC (Promega) was diluted into 50 mM TEAB pH 8 supplemented with 0.02% protease max (Promega) and added to the S-trap (Protifi). The column was placed inside a clean microcentrifuge tube and incubated at 47 °C in a

Thermomixer, without agitation, for one hour. After digestion peptides were sequentially eluted by spinning through 40 μ l of 50 mM TEAB pH 8, 40 μ l of 0.1% formic acid and 35 μ l of 50% aqueous acetonitrile with 0.2% formic acid. The resulting solution was combined and evaporated with a SpeedVac overnight.

Dried peptides were reconstituted in 0.1% formic acid and injected onto a reverse phase C18 IonOpticks 25 cm Aurora column using a nanoElute (Bruker Daltonics). The target on-column load was 200 ng total peptide per injection. The flow rate was 400 nl min⁻¹. Mobile phase A was 0.1% formic acid, 2% acetonitrile and 97.9% water; mobile phase B was 0.1% formic acid and 99.9% acetonitrile. The gradient was linear from 4% B to 36% B over 30 min. The mass spectrometer was a timsTOF Pro (Bruker Daltonics) set to acquire data in PASEF (parallel accumulation serial fragmentation) mode. The TIMS (trapped ion mobility spectrometry) ion accumulation time was set to 100 ms and precursors were pushed at 2.7 ms. The dynamic exclusion settings were set to exclude the same precursor masses for 0.4 min where charge states of 0–5 were allowed. For each sample three technical replicates were run and analysed independently.

Raw .d files from the timsTOF Pro were searched with Byonic v3.4 (Protein Metrics) using a custom FASTA database, which included the sequences of the human NTSR1 and human cysteine-free β arr1. The enzyme was trypsin (KR) and was set to fully specific, allowing for two missed cleavages and constrained to a false-discovery rate of less than 1%. The precursor mass tolerance was set to 20 ppm and the fragment ion tolerance at 40 ppm. A maximum of six common modifications and one rare modification were allowed per peptide, and the following modifications were included in the search: C carbamidomethyl (+57.021464) as a fixed modification; N-terminal acetyl (+42.010565) as rare 1; N and Q deamidated (+0.984016), as rare 1; M oxidation, (+15.994915) as common 1, ST phosphorylation (+79.966331) as common 4; Y phosphorylation (+79.966331) as rare 1; C palmitoyl (+238.229666) as rare 1. The resulting Byonic results were imported into Byologic v3.4-55 for quantitative analysis.

Lipidomics analysis of PtdIns(4,5)P₂

Measurements of diC8-PtdIns(4,5)P₂ were performed using a q-Exactive FT-mass spectrometer (Thermo) equipped with Vanquish Split Sampler FT uHPLC. Samples were diluted in 50% aqueous methanol and injected onto a Dionex Acclaim 120 C8 column (5 μ m particle size, 120 Å pore diameter, 2.1 mm internal diameter, 50 mm length). The column compartment was maintained at 29 °C. The flow rate was 0.3 ml min⁻¹. Mobile phase A was 1:1 (v/v%) MeOH/water, 5 mM ammonium acetate; mobile phase B was pure acetone. Each run was five minutes; the gradient used was as follows: 0–0.5 min 0% B, 0.5 to 2 min linear gradient from 0 to 50% B, 2 to 3.5 min hold at 50% B, 3.5 to 3.6 min from 50% B to 0% B and 3.6 to 5 min at 0% B to re-equilibrate the column. The first minute of the run was diverted to waste and minutes 1–4.8 were sent to the mass spectrometer for analysis. Detection on the q-Exactive was performed in negative mode between 200–2,000 m/z , using an acquisition target of 3×10^6 , maximum injection time of 100 ms at a resolution of 70,000 for MS and 17,500 for MS/MS (tandem MS) data. For MS/MS analysis, dynamic exclusion was set to 10 s and a global inclusion list was used to target the 745.2 m/z species. Automatic gain control target was set to 1×10^5 , maximum injection time was 100 ms for MS1 and 50 ms for MS2. A stepped collision energy from 25 eV to 35 eV (normalized) and an isolation window of 4 m/z was used for higher-energy collision-induced dissociation.

Crosslinking mass spectrometry

Excised SDS-PAGE slices corresponding to protein or crosslinked proteins of interest were destained and diced to 1-mm pieces before reduction with 5 mM dithiothreitol for 30 min at 55 °C, followed by alkylation with 10 mM acrylamide and finally digestion with trypsin/LysC (Promega). Digestion was performed at an estimated enzyme:substrate

ratio of 1:100 overnight at 37 °C. After proteolysis, the reaction was quenched with 1% formic acid. Peptides were extracted and dried using a SpeedVac before dissolving in 12 μ l of reconstitution buffer (2% acetonitrile with 0.1% formic acid); 3 μ l of this solution was injected onto the LC-MS. In a typical mass spectrometry experiment, an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) with liquid chromatography using a NanoAcquity UPLC (Waters Corporation) was used at a flow rate of 450 nl min⁻¹. Reverse-phase chromatographic separations consisted of mobile phase A (0.2% formic acid in water) and mobile phase B (0.2% formic acid in acetonitrile). The stationary phase consisted of fused silica columns—pulled and packed in-house—with an internal diameter of 100 μ m packed with Magic 1.8- μ m particle size, 120 Å pore diameter, UChrom C18 (nanoLCMS Solutions) at a length of around 25 cm. Peptides were directly injected onto the analytical column using a gradient (2–45% B, followed by a high-B wash) of length 80 min. The mass spectrometer was operated in a data-dependent fashion using either collision-induced decay (CID) fragmentation or higher-energy collisional dissociation (HCD)/electron-transfer dissociation (ETD) decision tree fragmentation for MS/MS spectra. For CID and ETD, MS/MS spectra were detected in the ion trap and for HCD, MS/MS spectra were detected in the Orbitrap. For data analysis, .RAW data files were processed using Byonic v3.2.0 (Protein Metrics) to identify peptides from a limited database of protein target sequences (as described in the section 'Phosphoproteomics experiments'). Peptides were assumed to be fully specific, allowing for common modifications (for example, oxidation of methionine), allowing for up to two missed cleavages and were constrained to a false-discovery rate of less than 2%. Mass accuracies for precursor and peptide fragments detected in the Orbitrap were held within 12 ppm, with 0.4-Da mass accuracies in the ion trap. Potential crosslinked peptides were validated using Byologic v3.4-55 (Protein Metrics). Search results were evaluated in Byologic and evaluated relative to the structural model of the NTSR1- β arr1 complex.

Fluorescence anisotropy measurements

BODIPY TMR phosphatidylinositol 4,5-bisphosphate (Echelon Biosciences) (henceforth BODIPY-PtdIns(4,5)P₂) was dissolved to a stock concentration of 1 mM in 50 mM HEPES pH 7.4 and used at a final concentration of 4 nM in the assay. For the arrestin measurements, a twofold dilution series of was made from a 311 μ M stock of β arr1(Δ CT), yielding fourteen samples with final concentrations ranging from 150 μ M to 0.02 μ M. For the NTSR1 measurements, a twofold dilution series was made from a 140 μ M stock of NTSR1, yielding fourteen samples with final concentrations ranging from 67 μ M to 0.01 μ M. For each, a control sample containing buffer only was included to measure the free anisotropy of BODIPY-PtdIns(4,5)P₂. After mixing the BODIPY-PtdIns(4,5)P₂ with the receptor, arrestin or buffer, samples were incubated for 1 h at room temperature before measurements. Samples were measured in five 20- μ l (arrestin) or three 10- μ l (receptor) replicates in a 384-well plate on a Tecan Infinite M1000 (Tecan Life Sciences), using an excitation wavelength of 530 nm, an emission wavelength of 573 nm and bandwidths of 5 nm. The obtained data was fit using 'One Site Total' nonlinear regression using GraphPad Prism 8.3.0.

NanoBiT β arr1 recruitment assay

Recruitment of β arr1 to NTSR1 was measured by a NanoBiT PPI assay⁵⁷. Human full-length β arr1 was N-terminally fused to a large fragment (LgBiT; forming Lg- β arr1) of the NanoBiT luciferase with a 15-amino-acid flexible linker (GGSGGGSGGSSGG). N-terminally Flag-epitope (DYKDDDDK)-tagged human NTSR1 was C-terminally fused to a small fragment (SmBiT; forming NTSR1-Sm) with the 15-amino-acid flexible linker. The Lg- β arr1 and the NTSR1-Sm constructs were inserted into a pCAGGS expression plasmid vector. HEK293A cells (Thermo Fisher Scientific, confirmed mycoplasma negative by manufacturer) were seeded in a 6-well culture plate at a concentration of 2×10^5 cells per ml (2 ml per well in DMEM (Nissui Pharmaceutical) supplemented with

10% fetal bovine serum (Gibco), glutamine, penicillin and streptomycin) 1 day before transfection. Transfection solution was prepared by combining 4 μ l (per well in a 6-well plate, hereafter) of polyethylenimine solution (Polysciences; 1 mg ml⁻¹) and a plasmid mixture consisting of 100 ng Lg- β arr1 and 500 ng NTSR1-Sm in 200 μ l of Opti-MEM (Thermo Fisher Scientific). After incubation for 1 day, transfected cells were collected with 0.5 mM EDTA-containing Dulbecco's PBS, centrifuged and suspended in 2 ml of HBSS containing 0.01% bovine serum albumin (BSA; fatty-acid-free grade; SERVA) and 5 mM HEPES (pH 7.4) (assay buffer). The cell suspension was dispensed in a white 96-well plate at a volume of 80 μ l per well and loaded with 20 μ l of 50 μ M coelenterazine (Carbosynth) diluted in the assay buffer. After 2-h incubation at room temperature, the plate was measured for baseline luminescence (Spectramax L, Molecular Devices) and 20 μ l of 6 \times neurotensin (Peptide Institute; final concentrations ranging from 0.1 nM to 10 μ M) diluted in the assay buffer or the assay buffer alone (vehicle) were manually added. The plate was read for 15 min with an interval of 20 s at room temperature. Luminescence counts were normalized to the initial count and arrestin association kinetics were calculated by fitting the normalized luminescence data to a one-phase association model built in Prism 8 software (GraphPad Prism). In cases in which luminescent kinetics were bell-shaped, data points over increasing luminescence were used for the fitting. Association speed was calculated by a formula of (plateau - Y_0)* K in which 'plateau' and Y_0 represent saturated and initial luminescent counts, respectively, and K denotes a rate constant in units that are the reciprocal of time (min). The resulting association speed data were fitted to a four-parameter sigmoidal concentration-response curve, from which pEC₅₀ values (negative logarithmic values of EC₅₀ values, in which EC₅₀ is the half-maximal effective concentration) and E_{\max} values ('top' - 'bottom') were used to calculate the mean and standard deviation.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The cryo-EM density map for the NTSR1- β arr1(Δ CT) complex has been deposited in the Electron Microscopy Data Bank under accession code EMD-20836. The coordinates for the model of NTSR1- β arr1(Δ CT) have been deposited in the PDB under accession number 6UP7. Proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository and can be accessed at <https://doi.org/10.6019/PXD016224>. The data supporting the findings of this study are available as Supplementary Information (Supplementary Tables 1–39 and Supplementary Figs. 2–5). All other data are available upon request to the corresponding authors.

46. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

47. Beyett, T. S. et al. Perturbation of the interactions of calmodulin with GRK5 using a natural product chemical probe. *Proc. Natl Acad. Sci. USA* **116**, 15895–15900 (2019).
48. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).
49. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
50. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
51. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
52. Heymann, J. B. & Belnap, D. M. Soft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
53. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
55. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
56. Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
57. Dixon, A. S. et al. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* **11**, 400–408 (2016).

Acknowledgements We thank M. Krawitzky and C. Adams for instrument and software access; R. Leib, F. Liu and M. Bern for discussions pertaining to mass spectrometry data analysis; G. Lam and B. Fitch for assistance with lipidomics analysis; K. Sato, Y. Sugamura and A. Inoue for plasmid construction and cell-based GPCR assays; and D. Mayer for suggestions on arrestin purification. This work was supported in part by National Institutes of Health grants R01NS028471 (B.K.K.), 1U19AI109662 (J.S.G.) and P30 CA124435 (for using the Stanford Cancer Institute Proteomics/Mass Spectrometry Shared Resource). Additional support to G.S. and B.K.K. was provided by the Mathers Foundation. B.K.K. is a Chan-Zuckerberg Biohub investigator. M.M. was supported by an American Heart Association postdoctoral fellowship (17POST33410958). J.J. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2318-18). A.I. was funded by the PRIME 18gm5910013 and the LEAP 18gm0010004 from the Japan Agency for Medical Research and Development (AMED) and KAKENHI 17K08264 from the Japan Society for the Promotion of Science (JSPS). H.E.K. was funded by KAKENHI 19H03163 from JSPS, The Naito Foundation, The Kurata Grants from The Hitachi Global Foundation, and Grant-in-Aid from the Tokyo Biochemical Research Foundation.

Author contributions W.H. initiated the project and performed GPCR screening to identify strong arrestin couplers with H.E.K. W.H. expressed and purified NTSR1, screened NTSR1 constructs and performed initial NTSR1- β arr1 complexing. M.M. established arrestin expression and purification, performed fluorescence polarization measurements, engineered pre-activated arrestin constructs and developed fluorescence and electron paramagnetic resonance reporters to guide the complex formation strategy. Q.Q. screened samples by negative-stain electron microscopy and cryo-EM, prepared grids, collected and processed cryo-EM data and generated the cryo-EM maps. J.J. expressed and purified GRK5, screened and optimized NTSR1 phosphorylation conditions, phosphorylated NTSR1 for cryo-EM studies, performed mass spectrometry experiments and analysed all mass spectrometry data. M.M. and J.J. established GRK5 expression and purification conditions and screened and optimized NTSR1- β arr1(Δ CT) complexing conditions. W.H. and J.J. chose crosslinkers to screen. W.H., M.M. and J.J. screened crosslinkers and optimized sample preparation for cryo-EM. W.H. and Q.Q. build the NTSR1- β arr1(Δ CT) model, with contributions from J.J. and M.J.R. K.C.N. performed lipidomics measurements, which were overseen by J.J. and J.S.G. A.I. performed NanoBIT experiments. M.M., J.J., G.S. and B.K.K. wrote the paper with input from W.H. and Q.Q. B.K.K. and G.S. supervised the project.

Competing interests B.K.K. is a co-founder of and consultant for ConfometRx.

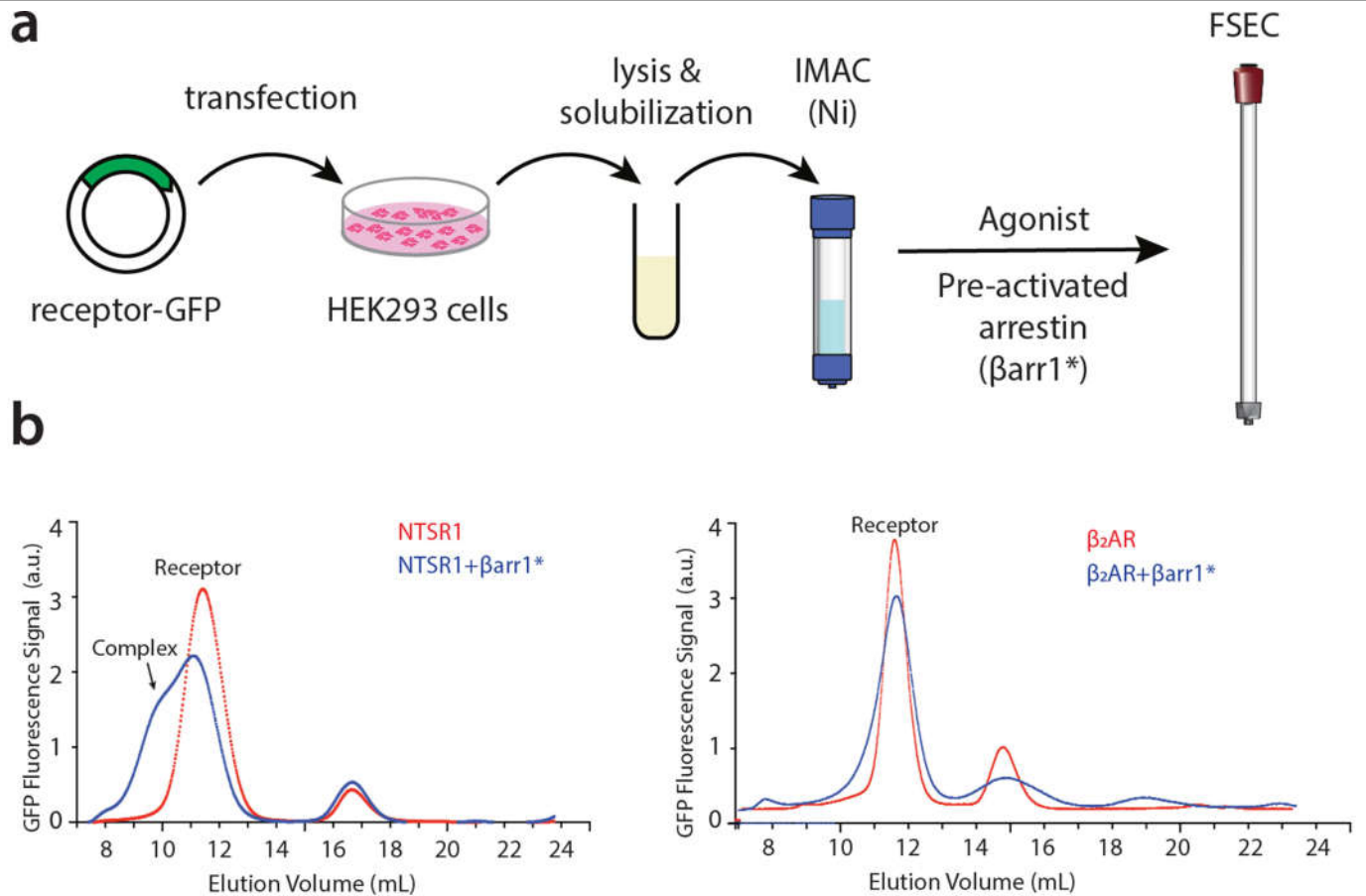
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1953-1>.

Correspondence and requests for materials should be addressed to G.S. or B.K.K.

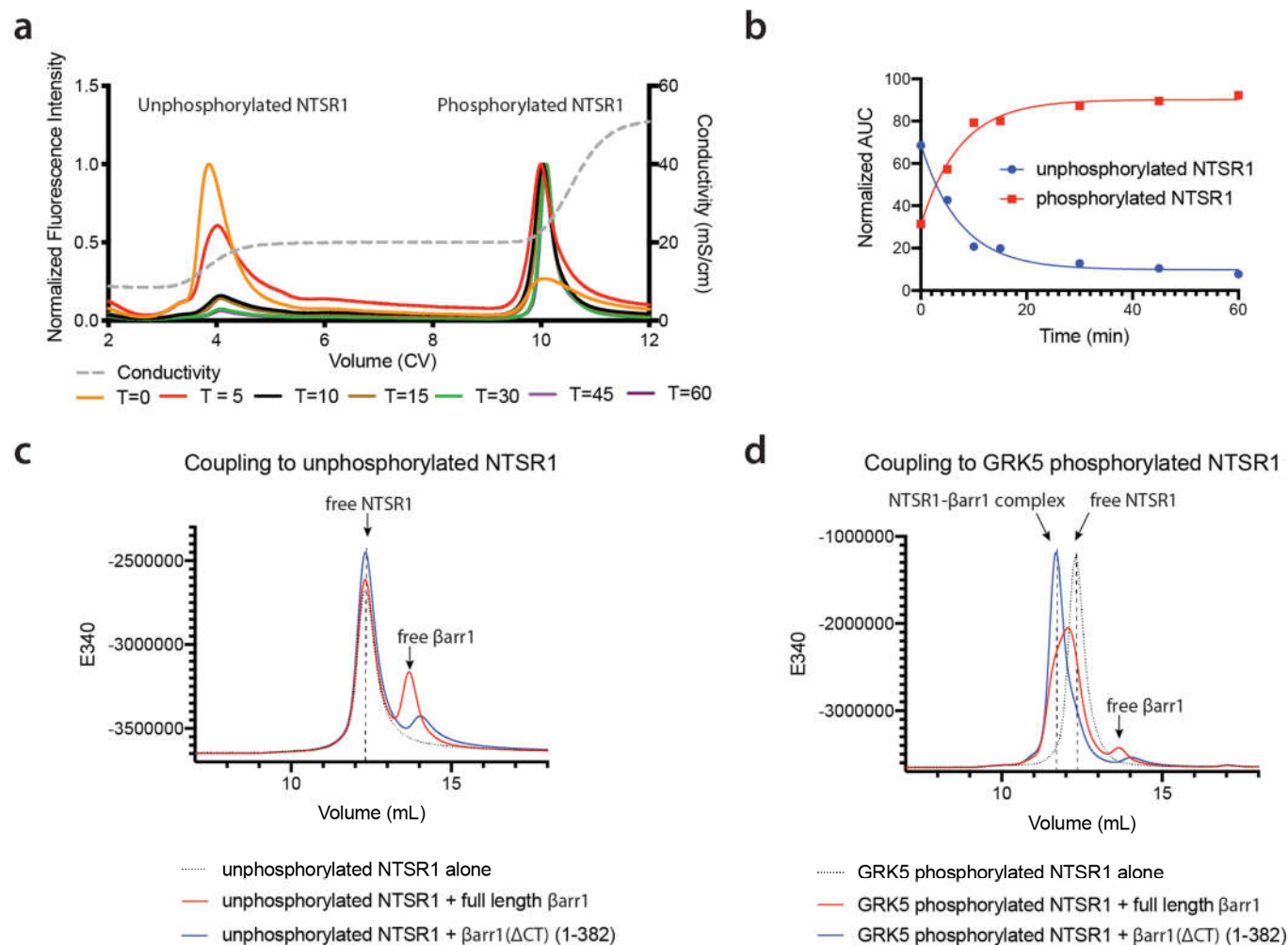
Peer review information Nature thanks Oliver Clarke, Martin Lohse and John Tesmer for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



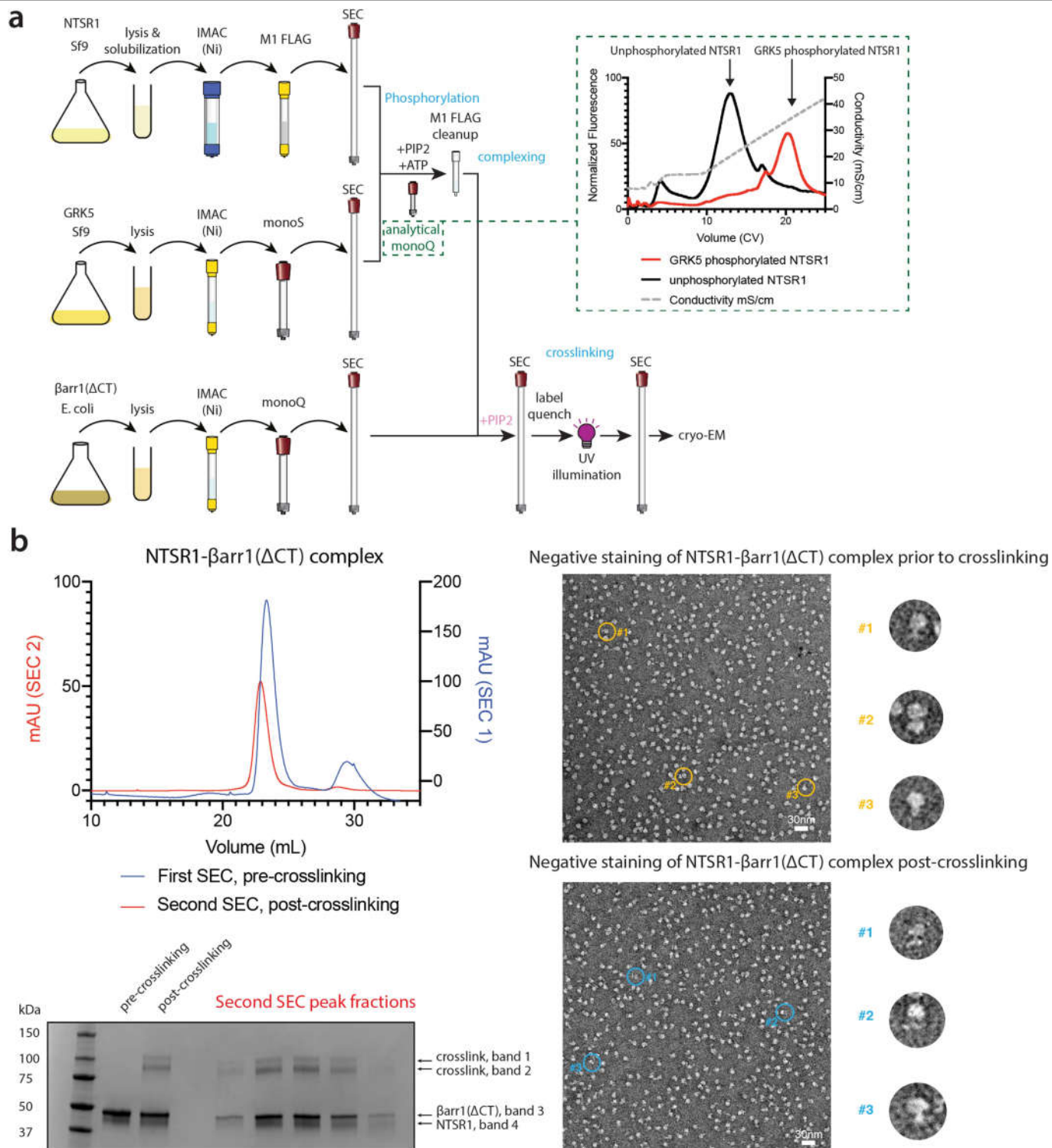
Extended Data Fig. 1 | Screening for GPCR- β arr1 complexes. A fluorescence size-exclusion chromatography (FSEC) approach was used to identify GPCRs that display strong coupling to β arr1 in vitro. **a**, Overview of the screening strategy. Small-scale transfection and purification of candidate GFP-tagged GPCRs was followed by complexing with activated β arr1, denoted by β arr1*. β arr1 was activated in situ by mixing with Fab30 and V2Rpp (Methods) before

being added (in excess) to the agonist-bound receptor. Complex formation was assessed by FSEC. **b**, Representative FSEC traces. The trace of NTSR1 + β arr1* (left) is notable in that it shows a shoulder peak, which indicates substantial complex formation, whereas the trace for β_2 adrenergic receptor (β_2 AR, right) + β arr1* shows little or no complex formation. Experiments were performed independently twice, with similar results.



Extended Data Fig. 2 | Phosphorylation of NTSR1 is crucial for the coupling of βarr1. **a**, The phosphorylation state of NTSR1 was assessed using ion-exchange chromatography. Chromatograms of aliquots taken at various time points (T , shown in minutes) are overlaid, and the relative amounts of unphosphorylated and phosphorylated species were measured using an optimized stepped elution profile. This experiment was performed once. **b**, Graphical representation of normalized relative ratio of phosphorylated and unphosphorylated receptor from data shown in **a**. **c**, d, FSEC chromatogram for

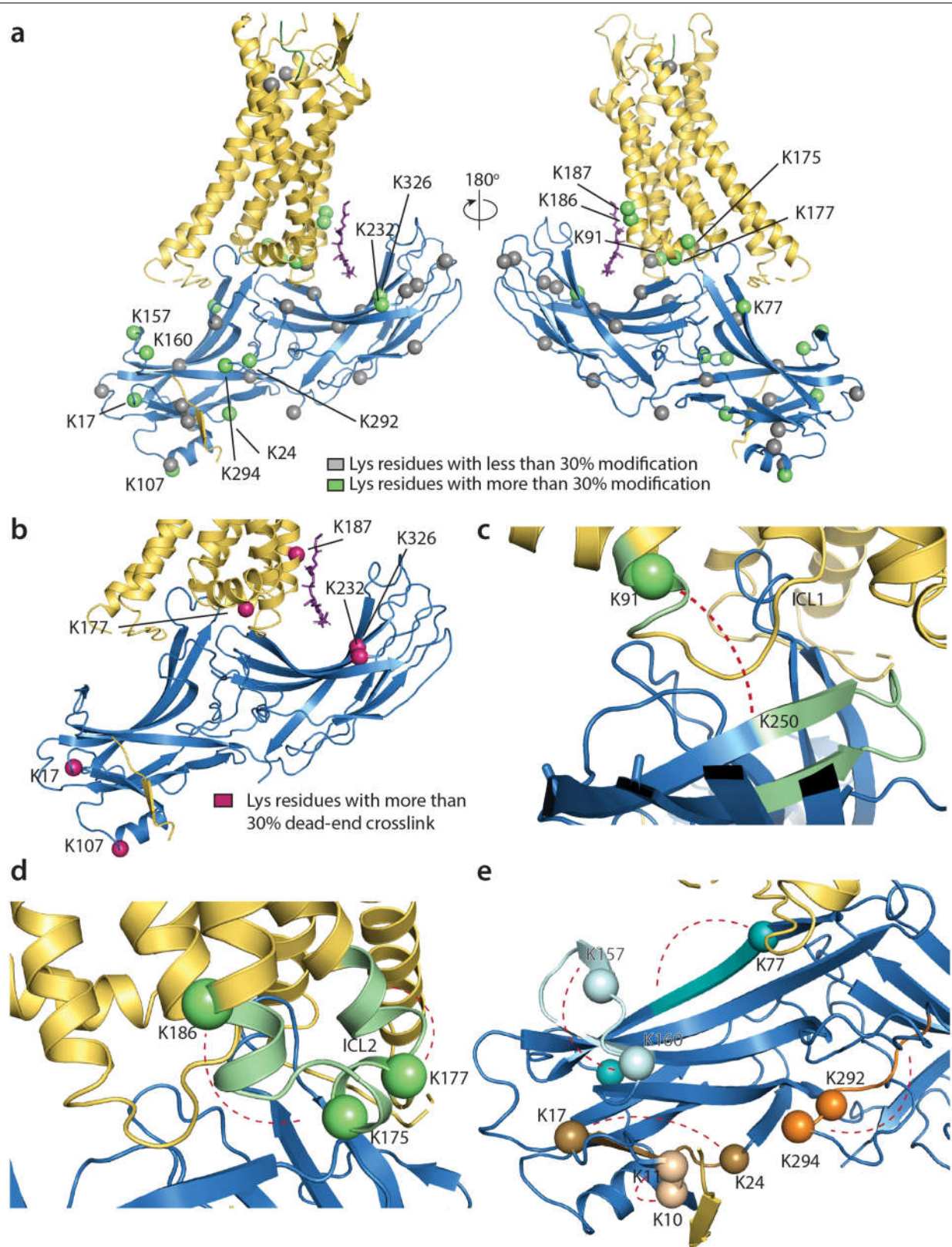
screening βarr1 constructs for forming a stable complex with either unphosphorylated (**c**) or GRK5-phosphorylated (**d**) NTSR1. Elution volumes for free receptor, free βarr1 and the receptor-βarr1 complex are shown. The unphosphorylated receptor does not form a complex with any of the βarr1 constructs tested, whereas the phosphorylated receptor couples moderately to full-length βarr1 and strongly to the pre-activated βarr1(ΔCT) construct. Experiments were performed independently twice, with similar results.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Overview of sample preparation and representative cryo-EM preparation. a, Workflow for the assembly of the NTSR1- β arr1(Δ CT) complex. Receptor, GRK5 and β arr1(Δ CT) were purified separately. Phosphorylation of NTSR1 by GRK5 in the presence of diC8-PtdIns(4,5) P_2 was monitored by analytical ion-exchange chromatography (shown in inset as gradient elution from a representative preparation; a stepped elution ion-exchange chromatography run is shown in Extended Data Fig. 2a), followed by M1 affinity purification. The phosphorylated receptor was combined with β arr1(Δ CT) and additional diC8-PtdIns(4,5) P_2 and separated by size-exclusion chromatography (SEC) to isolate the NTSR1- β arr1(Δ CT) complex. The complex was labelled with sulfo-LC-SDA and irradiated with UV light. After a second round of SEC to re-isolate the complex, the sample was concentrated and used for cryo-EM. **b,** Representative sample preparation: phosphorylated NTSR1

was mixed with β arr1(Δ CT) and purified by SEC (blue curve, SEC1). Complex fractions were combined and treated with crosslinker, quenched, then UV-irradiated. Lanes 2 and 3 of the SDS-PAGE gel show pooled samples before and after crosslinking and UV irradiation; about 25% of the sample is crosslinked (based on densitometry, top two bands relative to total). The crosslinked sample was re-run on SEC (red curve, SEC2) and SDS-PAGE of peak fractions is shown. Peak fractions were combined, concentrated and used for cryo-EM. Representative negative-stain electron microscopy images of the NTSR1- β arr1(Δ CT) complex show similar homogeneity, pre- and post-crosslinking. Enlarged views of rotated representative particles (denoted by yellow and blue circles) are shown. Four sample preparations, which all gave similar results by SEC, SDS-PAGE and negative-stain analysis, were used to generate the cryo-EM dataset.

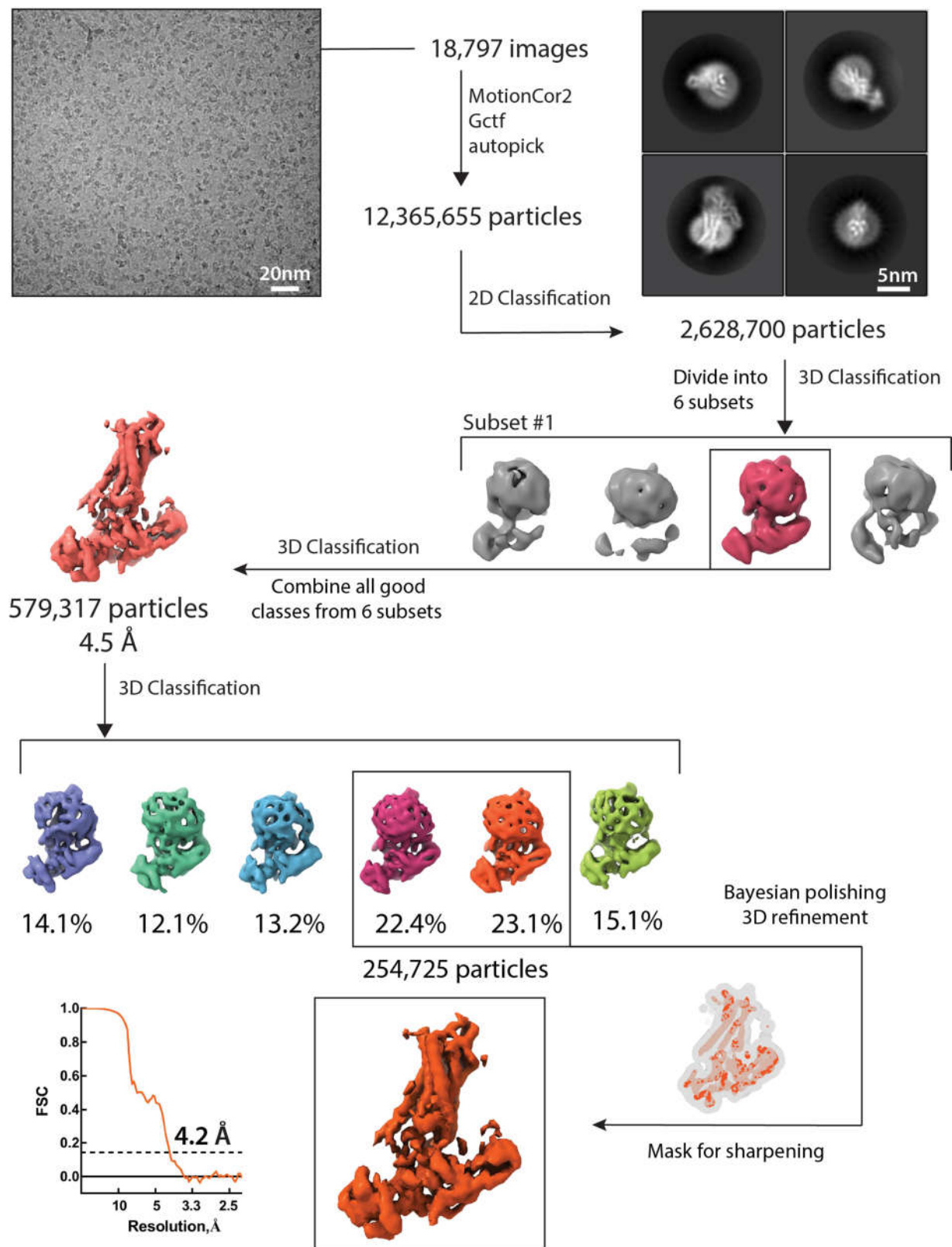


Extended Data Fig. 4 | See next page for caption.

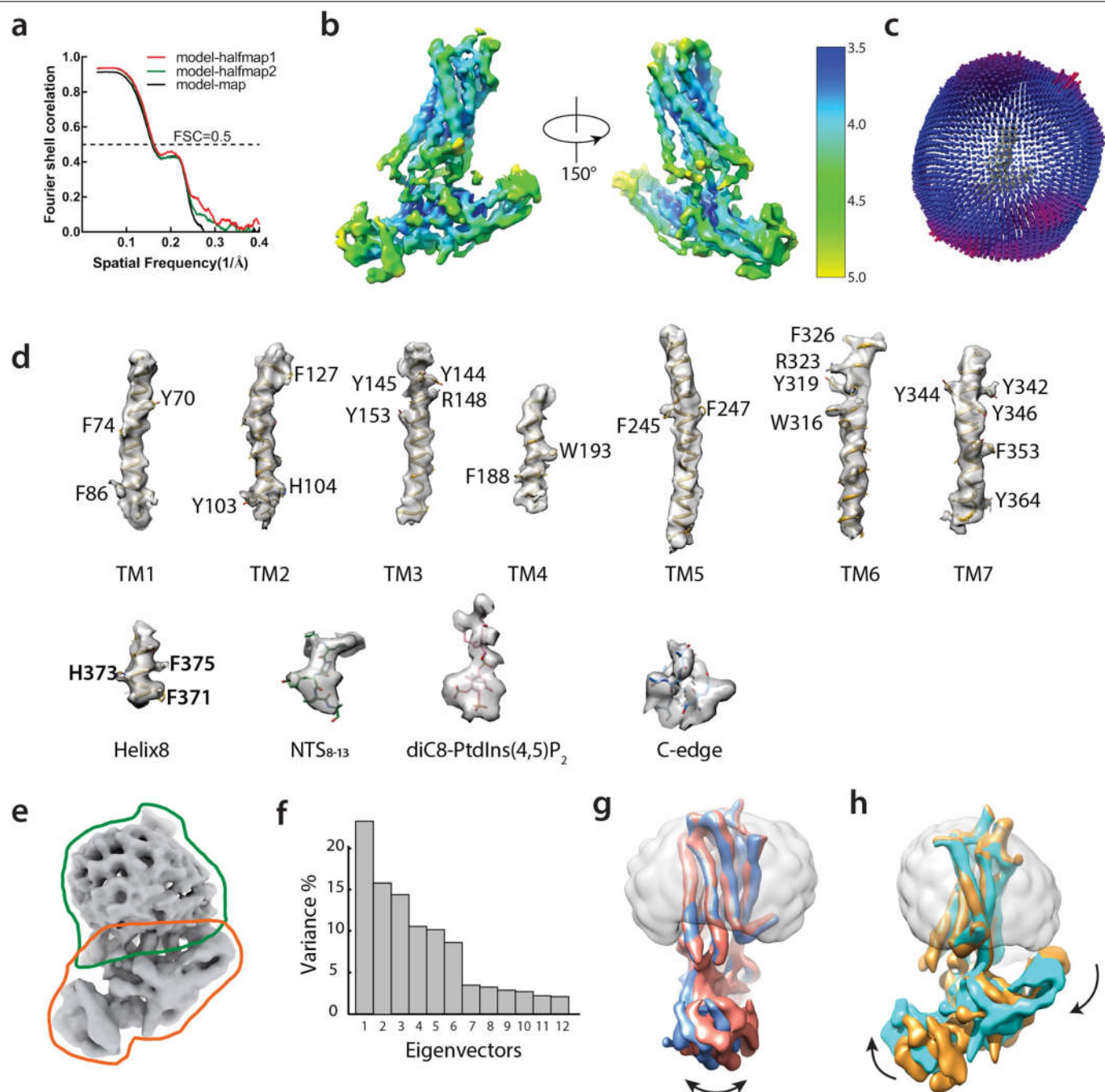
Extended Data Fig. 4 | Mass spectrometry analysis of crosslinks in the

NTSR1- β arr1(Δ CT) complex. a, Lysine residues modified within the NTSR1- β arr1(Δ CT) complex after treatment with sulfo-LC-SDA were identified by mass spectrometry. Although mass spectrometry data were collected for each band observed by SDS-PAGE separately (Extended Data Fig. 3b), data were combined for analysis as the overall sample was imaged by cryo-EM. For each lysine residue in the complex the degree of modification was approximated as the percentage of times a lysine was found to be modified relative to the total number of times that lysine is observed across all peptides. The C α for lysine residues that were observed to be modified less than 30% of the time are shown as grey spheres, whereas those found to be modified more than 30% of the time are shown as green spheres. A tabulated analysis is presented in Supplementary Tables 1-7. **b,** Of the modified lysine residues, several mainly formed dead-end crosslinks (reaction of the activated diazirene with water);

those with more than 30% of the lysine residues resulting in dead-end crosslinks are shown as maroon spheres (C α). **c,** Only one intermolecular crosslink is observed that is consistent with the structural model obtained from cryo-EM. This crosslink was localized to ICL1 (K91) and a β -strand (shown in green) in the arrestin C-lobe. **d,** Several intramolecular crosslinks within ICL2 and the adjacent helices are observed. Reactive lysine residues are shown as green spheres (for C α), and crosslinks were localized to the region shown in green. **e,** Multiple intramolecular crosslinks within arrestin were observed. Each reactive lysine residue (shown as a C α sphere) that forms an intrapeptide crosslink, or an intramolecular (within arrestin) crosslink, is coloured to match the region in which the observed crosslink could be localized. Each band was analysed once by mass spectrometry from a representative cryo-EM preparation.

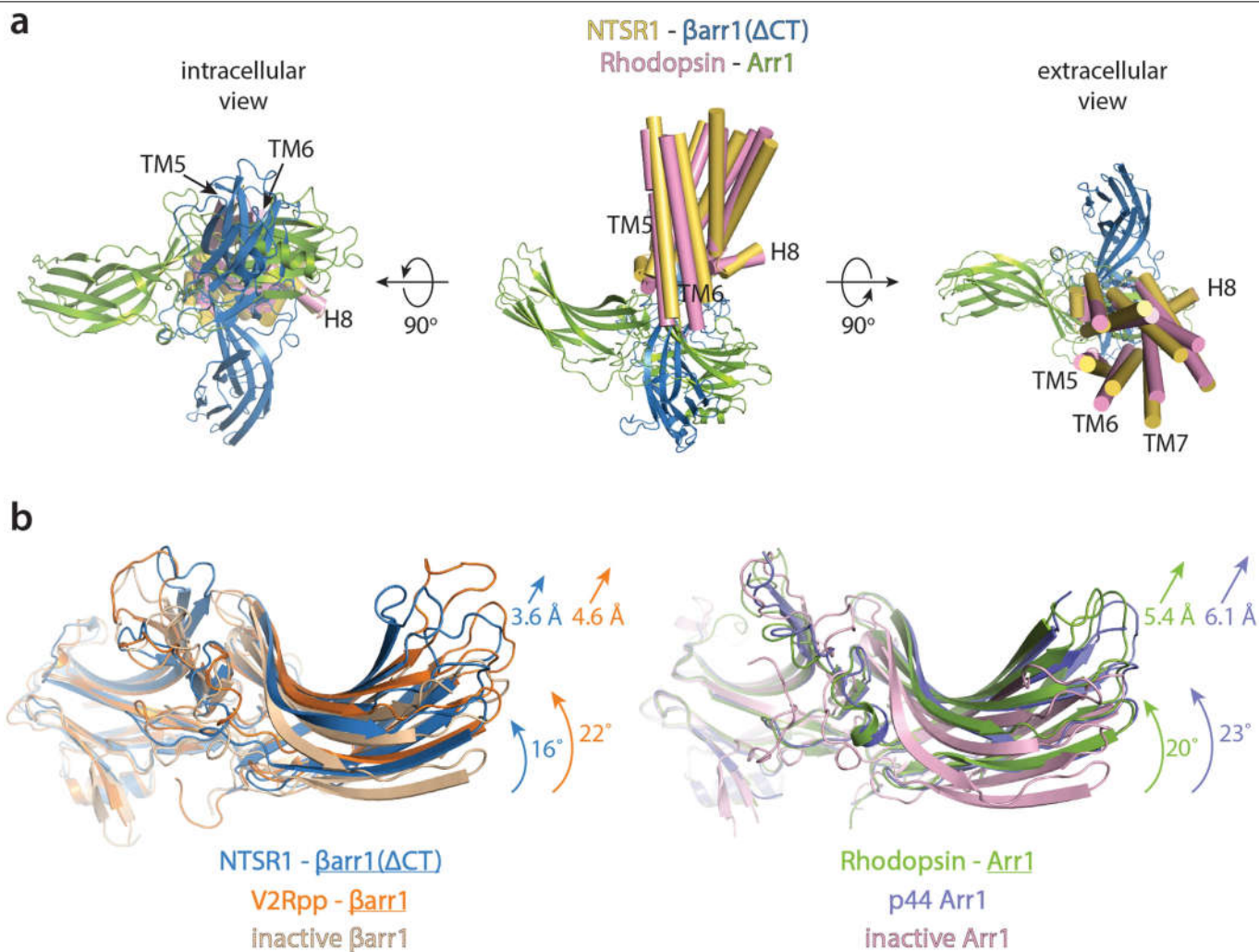


Extended Data Fig. 5 | Cryo-EM workflow. Flow chart of cryo-EM data analysis. Global resolution was determined by FSC with a cut-off of 0.143. The mask used for sharpening was also used for global resolution determination.



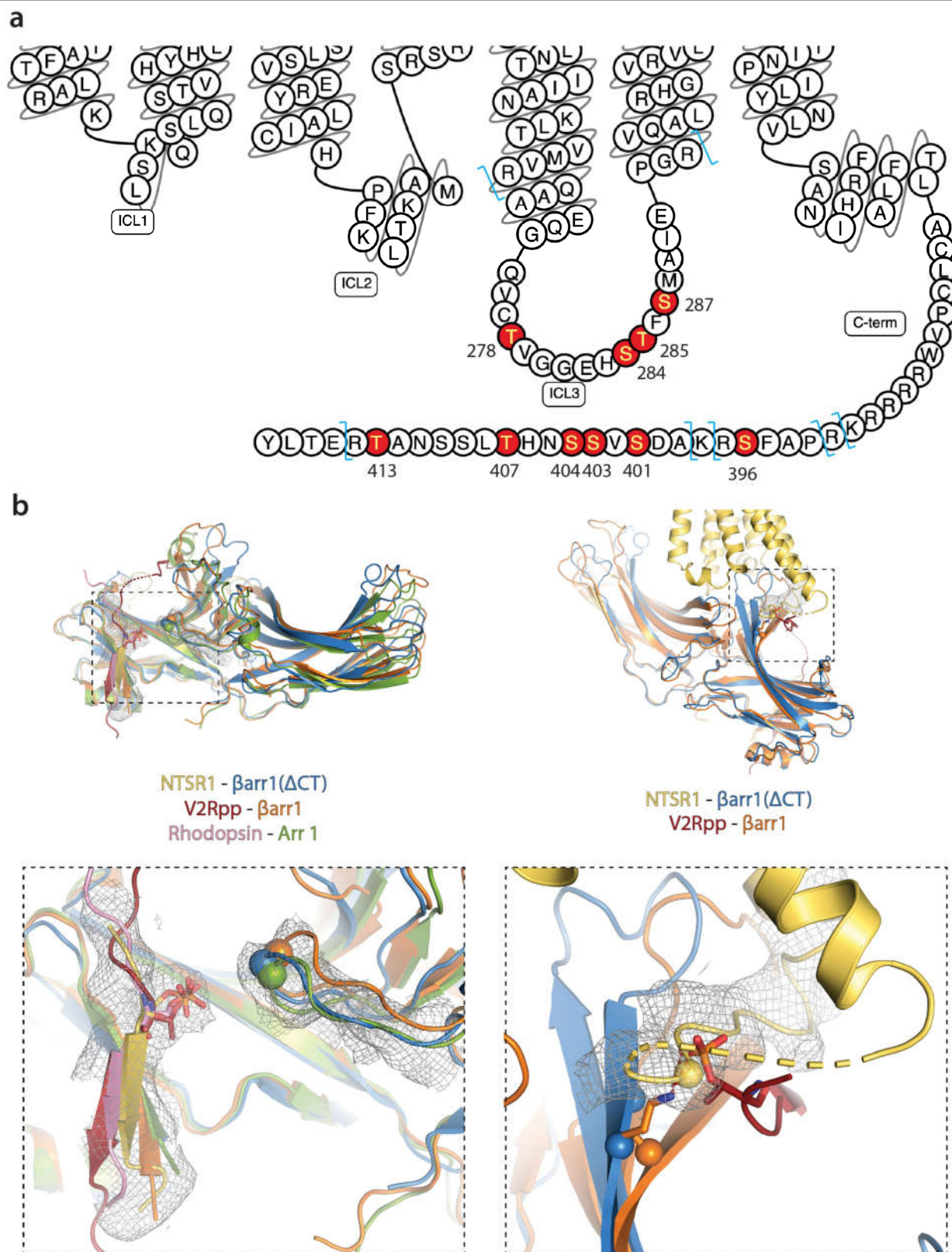
Extended Data Fig. 6 | Cryo-EM map resolution and model validation, with multi-body analysis revealing structural heterogeneity within the NTSR1- β arr1(Δ CT) complex. **a, Cross validation between the map and the model. The model was refined against one half map after displacement of atoms by 0.2 Å, and FSC curves were calculated between this model and the working half1 map (red), the free half2 map (green) and the final cryo-EM map (full dataset, black) by Mtriage implemented in PHENIX. **b**, Local resolution of the final 4.2 Å map was estimated by Bsoft. **c**, Euler angle distribution of the particle set used in the final map. **d**, Representative sections of the model with accompanying regions**

of density from the electron microscopy map. **e**, Masks used for multi-body refinement. **f**, The contributions of each of the twelve eigenvectors (numbered along the x-axis) to the variance of the overall final map. **g**, Maps corresponding to the second and ninth components of the first eigenvector (**f**) are aligned, showing swing-like motion of β arr1(Δ CT) with respect to NTSR1. **h**, Maps corresponding to first and tenth components of the second eigenvector (**f**) are superimposed, indicating the tilt-like motion of β arr1(Δ CT) with respect to NTSR1.



Extended Data Fig. 7 | Arrestin orientation and interdomain twist. a, Additional views of Fig. 1c, showing an overlay between NTSR1- β arr1(Δ CT) and Rho-Arr1: cytoplasmic (intracellular) view (left), side view (middle) and extracellular view (right). **b,** Compared with the inactive structure, β arr1(Δ CT) in complex with NTSR1 displays an additional interdomain twist between its N-lobe and its C-lobe, which is a hallmark of activation. Comparison of the interdomain twist between NTSR1- β arr1(Δ CT) (blue, present structure) and V2Rpp- β arr1 (orange, PDB: 4JQI) relative to inactive β arr1 (wheat, PDB: 1G4M).

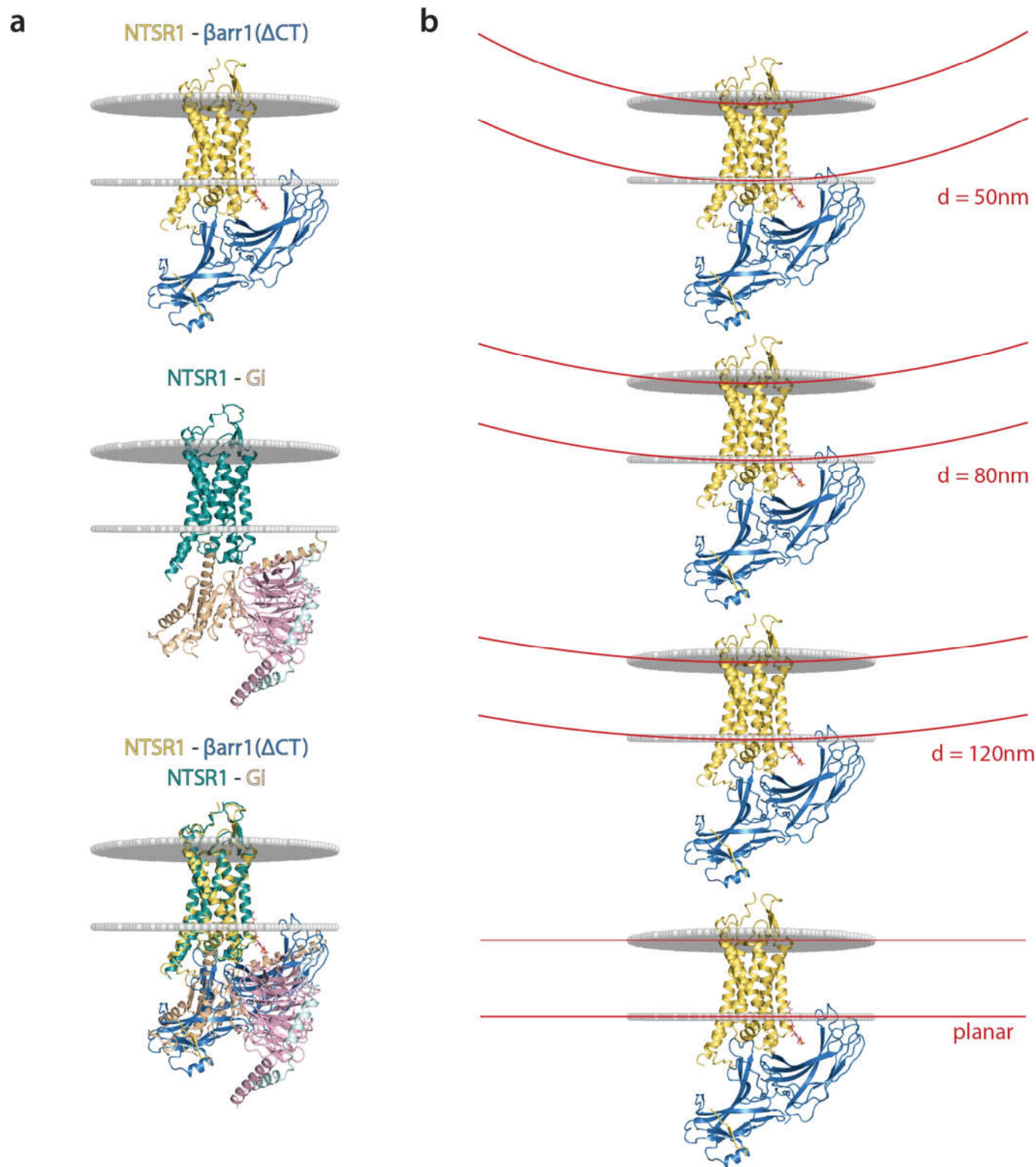
A smaller interdomain twist and displacement between respective C-lobes is observed for NTSR1-bound β arr1 (16°, 3.6 Å) compared with V2Rpp-bound β arr1 (22°, 4.6 Å), indicated by arrows drawn to relative scale. A similar comparison for rhodopsin-bound arrestin-1 (green, PDB: 5W0P) and pre-activated p44 arrestin-1 (purple, PDB: 4J2Q) relative to inactive arrestin-1 (pink, PDB: 1CF1) yields interdomain twist and displacement values of 20°, 5.4 Å and 23°, 6.1 Å, respectively.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Identification of GRK5 phosphorylation sites in human NTSR1 and comparison of phosphate-specific contacts in NTSR1- β arr1(Δ CT) to those in Rho-Arr1 and V2Rpp- β arr1. **a**, Snake diagram of human NTSR1 showing identified phosphosites in ICL3 and the receptor C terminus. The trypsin/LysC digestion sites are shown in light blue. All four phosphosites for ICL3 are in a single peptide beginning with QAA at the end of TM5 and ending at PGR at the beginning of TM6. Phosphosites in the C terminus occur on two peptides, one containing S396, RPAFSR, as well as several missed cleavage variants; and one containing S401, S403, S404, T407 and T413. Tabulated fragment ions are presented in Supplementary Tables 8–39 and spectra are shown in Supplementary Figs. 2–5. Any post translational modifications outside of these two discussed regions are not shown for clarity.

These experiments were performed independently twice, and analysed by two or three technical replicates each time, all with similar results. **b**, Overlay of arrestin structures, in which binding of phosphorylated receptor residues contributes to stabilizing the active state of arrestin. Left, equivalent phosphoresidues on different receptor C-tails (shown in stick representation and coloured accordingly: NTSR1, yellow; V2Rpp, red; rhodopsin, pink) contribute to arrestin activation by interacting with a lysine residue on the gate loop (K294, shown as α spheres), thereby breaking the polar core. Right, phosphoresidue pT347 in the V2Rpp interacts with residue K77 at the base of the finger loop. A similarly positioned phosphoresidue in the ICL3 of NTSR1 (pS287) probably interacts with arrestin residues R76/K77 in a similar manner.



Extended Data Fig. 9 | Membrane context and curvature for NTSR1- β arr1(Δ CT). **a**, Comparison of the structures of NTSR1- β arr1(Δ CT) and NTSR1-cGi, for which membrane-embedded regions were computed using the Orientations of Proteins in Membranes server (<https://opm.phar.umich.edu/>). The membrane is indicated by grey spheres to show the extent to which the C-edge would insert into the membrane. **b**, The extent of arrestin tilt observed

in the NTSR1- β arr1(Δ CT) structure could also be relevant to the increased membrane curvature observed in endocytic vesicles. As an example, several membrane bilayers are drawn (in red) to scale, comparing the degree of curvature for vesicles ranging from 120 nm to 50 nm in diameter relative to a planar bilayer.

Cryo-EM data collection, refinement and validation statistics

	NTSR1- β arr1(Δ CT) (EMDB-20836) (PDB 6UP7)
Data collection and processing	
Magnification	130,000
Voltage (kV)	300
Electron exposure (e-/Å ²)	56
Defocus range (μm)	-1.0~-2.0
Pixel size (Å)	1.06
Symmetry imposed	C1
Initial particle images (no.)	12,365,655
Final particle images (no.)	254,725
Map resolution (Å)	4.2
FSC threshold	0.143
Map resolution range (Å)	3.5~5.0
Refinement	
Initial model used (PDB code)	4GRV,4JQI
Map sharpening <i>B</i> factor (Å ²)	-164
Model composition	
Non-hydrogen atoms	4166
Protein residues	685
Ligands	1
<i>B</i> factors (Å ²)	
Protein	69.94
Ligand	130.87
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.362
Validation	
MolProbity score	2.09
Clashscore	8.44
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	86.81
Allowed (%)	13.19
Disallowed (%)	0

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Automated data collection on the Titan Krios was performed using serialEM 3.7.10

Data analysis

The following software was used in this study: Bsoft 2.0.6, Molprobit 4.5, MotionCor2, gCTF1.06, RELION 3.0, UCSF Chimera 1.14, UCSF ChimeraX 0.9, Coot 0.8.9, Phenix 1.16, Graphpad Prism 8.3.0, Byonic v3.4 and Byologic v.3.4-55 (Protein Metrics Inc.), Excalibur v4.1, and DataAnalysis v5.2 (Bruker Daltonics).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability statement included, all data has been deposited

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not predetermined by statistical methods. For cryoEM data, sample sizes were determined by availability of microscope. Cryo-EM data was collected until we were able to refine a high-resolution structure that allowed us to obtain a high-resolution reconstruction within the confines of limited microscope time.
Data exclusions	No data was systematically excluded. MS search results were manually inspected to remove false positive IDs. The process of generating 3D maps from cryo-EM particles involves sorting for particles that are damaged, have low signal, or are in minority conformations that are unlikely to refine correctly. This is implemented in Relion 2.1
Replication	All attempts at replication of biochemical and signaling assays succeeded. For cryoEM, data was collected over several rounds of data collection using multiple samples prepared on different days and with different input materials. NanoBiT G-protein dissociation assay shows mean +/- SD from 5 independent experiments performed in duplicate (technical replicates of the biological replicates).
Randomization	No randomization was attempted or needed. This was not a clinical trial or animal study that is dependent on randomization. All variables could be controlled.
Blinding	No blinding was attempted or needed. There was no group allocation performed in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Sf9, Expression Systems, Cat 94-001S. HEK293 Cells (ATCC CRL 1573) HEK293A cells, ThermoFisher Scientific, Cat R70507.
Authentication	Cell lines are maintained by the supplier. No additional authentication was performed by the authors of this study.
Mycoplasma contamination	Cell lines are tested by manufacturer for contamination and no additional testing was performed by the authors of this study.
Commonly misidentified lines (See ICLAC register)	None used.

Author Correction: Bile acid metabolites control T_H17 and T_{reg} cell differentiation

<https://doi.org/10.1038/s41586-020-2030-5>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1785-z>

Published online 27 November 2019



Check for updates

Saiyu Hang, Donggi Paik, Lina Yao, Eunha Kim, Jamma Trinath, Jingping Lu, Soyoung Ha, Brandon N. Nelson, Samantha P. Kelly, Lin Wu, Ye Zheng, Randy S. Longman, Fraydoon Rastinejad, A. Sloan Devlin, Michael R. Krout, Michael A. Fischbach, Dan R. Littman & Jun R. Huh

In this Article, author Jamma Trinath was incorrectly listed as Trinath Jamma. This error has been corrected online.

Publisher Correction: Modular and tunable biological feedback control using a de novo protein switch

<https://doi.org/10.1038/s41586-020-2046-x>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1425-7>

Published online 24 July 2019

 Check for updates

Andrew H. Ng, Taylor H. Nguyen, Mariana Gómez-Schiavon, Galen Dods, Robert A. Langan, Scott E. Boyken, Jennifer A. Samson, Lucas M. Waldburger, John E. Dueber, David Baker & Hana El-Samad

In Fig. 3f of this Letter, owing to an error during the production process, the y-axis label of the top graph was inadvertently shown as ‘YFP fluorescence (AU)’ instead of ‘RFP fluorescence (AU)’. This error has been corrected online.

Work

Your
story

Send your careers story
to: naturecareerseditor@nature.com



JARGON SHUTS READERS OUT

Non-scientists feel confused by technical language – even if it's defined. **By Chris Woolston**

Overly technical language in science articles doesn't just muddy the waters for non-experts – it can alienate readers, potentially shutting them out from scientific discussion and knowledge. That's the conclusion of a study published in the *Journal of Language and Social Psychology*¹, and it applies to general-interest articles just as much as to scientific papers.

"When we have a hard time conceptualizing information, we become really scared of it," says lead author Hillary Shulman, a communication researcher at the Ohio State University in Columbus. Scientists can create "unnecessary barriers" with words, she says.

The study involved 650 members of the general public who read paragraphs on three topics: self-driving cars, robotic surgery and 3D bio-printing. The paragraphs were either

laden with jargon terms, such as "remote ergonomic console", or written with words that are familiar to most readers, such as "separate control panel". Shulman and her co-authors wrote the texts using jargon gathered from articles and websites aimed at all readers, not from scientific journals or technical manuals.

After reading the passages, the study participants rated their experience in a series of questionnaires. Those who read jargon-filled paragraphs were more likely to say that they had difficulty understanding the language and the information. They were also significantly more likely to say that they weren't good at science, and less likely to say that they would seek out information on the topic in the future.

Some of the participants who read the jargon-heavy text received links to definitions of technical terms, but that didn't reduce

their frustrations or enhance their feelings of understanding. "We found that people didn't use the links," Shulman says. Instead of trying to define technical language when communicating with non-experts, she says, scientists would do better to avoid any such terms.

Scientists can learn to cut back on their use of technical language when talking to people who are not researchers, says Ayelet Baram-Tsabari, a science-communication researcher at the Technion Israel Institute of Technology in Haifa. In 2017, she helped to develop the De-Jargonizer, an online tool that assesses and scores the accessibility of text².

Baram-Tsabari also co-authored a January study in *PLoS ONE*³ showing that scientists with media training can write articles that are just as engaging as pieces written by professional journalists. "Avoiding jargon is a fundamental part of that, but it's not the whole story," she says. To really connect with the public, she recommends that scientists tell a story that's relevant to the audience.

Members of the public aren't the only ones who can be turned off by jargon, Shulman says. Students can be, too. "I teach a class with 400 undergrads," she says. "When you're training people, you can introduce jargon with a little more sensitivity. You're trying to invite them into the environment."

Of course, technical words still have an important function in science. Shulman's paper is itself loaded with terms such as 'metacognition' and 'self-schema'. "The irony of that is not lost on me," she says. "When it comes to scientific literature, you can't get anything published unless it's full of jargon. Scientists want to speak to other scientists in the most precise way possible."

Baram-Tsabari says that one of her graduate students has gathered anecdotal evidence that some female researchers feel especially pressured to use heavily technical words and phrases. "They say, 'People don't take me seriously because I'm a woman. If I used accessible language, it would be bad for my career.'" Similarly, Shulman has noticed that early-career researchers tend to lean on technical language to show that they belong in the community. "I see it a lot in graduate students," she says.

Chris Woolston is a freelance writer in Billings, Montana.

1. Shulman, H. C., Dixon, G. N., Bullock, O. M. & Colón Amill, D. J. *Lang. Soc. Psychol.* <https://doi.org/10.1177/0261927X20902177> (2020).
2. Rakedzon, T., Segev, E., Chapnik, N., Yosef, R. & Baram-Tsabari, A. *PLoS ONE* **12**, e0181742 (2017).
3. Barel-Ben David, Y., Garty, E. S. & Baram-Tsabari, A. *PLoS ONE* **15**, e0222250 (2020).

INTERCONTINENTAL CULTURE CLASH

A move from Germany to Mexico was a crash course in acclimatizing to a new research environment. **By Deb Raj Aryal**

In 2010, after obtaining my master's degree in Germany, I moved to Mexico to pursue my PhD in ecology and sustainable development. In both places, far away from my home country of Nepal, I worked with scientists and other colleagues from different cultures. This experience greatly improved my interpersonal skills and broadened my view of the world of research.

After I completed my PhD programme in 2015, I started a research fellowship at Chiapas Autonomous University in Mexico. My work focused on the environmental impact of agriculture and livestock production in the tropics and subtropics. It was part of a programme set up by the Mexican National Science and Technology Council (CONACYT) to attract foreign researchers and to encourage Mexican scientists based abroad to return to the country. The programme ran from June 2014 to December 2018 and attracted more than 1,300 early-career researchers.

I was conscious that the arrival of so many scientists over four years in a country with limited research funding might make for an uncomfortable situation for the established scientists already working there. CONACYT research fellows would be absorbed as staff scientists or faculty members into their host institutes within ten years, although this was not obligatory. As someone who was part of this large cohort of researchers, I have some advice on how to adapt to a new working environment to the benefit of both yourself and your colleagues.

Be humble

It is an advantage to be an early-career scientist with a lot of energy and enthusiasm for carrying out cutting-edge research with the latest tools and methods – but it's important to not come across as arrogant. An over-confident or pompous attitude can result in colleagues not wanting to cooperate or collaborate with you. At the beginning of their careers, researchers are most successful when they are humble and value the work of others, in my view.

Respect differences

Differences in work styles or cultures can feel more significant if you come from a different continent. Generally, new postdocs



RICHARD DRURY/GETTY

or principal investigators at an institute should not aim to change the work culture of colleagues who have been at that place for years. Unless there is unfair treatment or harmful behaviour, try instead to adapt to the new culture. There could be moments in which you feel uncomfortable with the existing dynamics and working styles, but

“It’s okay that you don’t know the local norms and regulations at first.”

it's important to accept that everyone has their own way of working. Constant and open communication regarding your worries and anxieties can help to create a shared sense of well-being in cross-cultural research labs.

Handle questions calmly

Scientists with many years of experience in your field might know more than you do. If you are asked a question that you feel is hostile, don't take it as a personal attack on the integrity of your work. Instead, thank the researcher who asked it, but be honest and say that you're still learning. As an

early-career researcher, you are not expected to know everything. If you find yourself being compared with others, don't try to prove that you are better. Your research should speak for itself. Sometimes, one group of researchers might disagree with another on how to tackle a project or share lab equipment, for example. Taking the side of the one group by criticizing the other won't help scientists who have just started a job at the institute.

Believe in yourself

Have faith in your capability to adjust, to work well with others and to deliver on the research responsibilities you are assigned. It's okay that you don't know the local norms and regulations at first. Use common sense and think about how you might want to be treated. If a question comes up that you don't know the answer to, ask someone; many people are willing to help newcomers. It is important to have a strong and lasting motivation for your career in science. It's fine if your first attempt at your research project doesn't succeed – but you must persevere when you fail.

Deb Raj Aryal is a CONACYT research fellow at the Autonomous University of Chiapas in Mexico.
e-mail: drajar@conacyt.mx



Where I work Rose A. Marks

I study how plants adapt to some of the most extreme conditions on Earth. My particular focus is the South African 'resurrection' plant (*Myrothamnus flabellifolia*), which can survive for years in a completely desiccated state. Understanding what makes these plants so resilient could help us to develop crops that will survive drought.

I often travel to the Buffelskloof Nature Reserve, in the Mpumalanga province of South Africa, to measure and take samples of *M. flabellifolia* that grow there, on a 100-metre quartzite cliff. In this picture, where I'm about about to abseil down it, I'm recording the plants' height, sex, number of flowers, hydration status and soil depth. I aim to understand how their structure, life history and drought tolerance change over time, and with distance up the cliff face.

I'd been rock climbing for more than a decade before I began researching these plants. I never had any formal training in abseiling, but friends and mentors taught me the basics, and I learnt through experience. I've broken both ankles and once fell

15 metres before the rope caught me.

At some point, I realized that I could use my abseiling skills to reach rarely studied plant communities in extreme habitats. I've found that the flowers of *M. flabellifolia* are also desiccation-tolerant and that the plant dries at different rates on the cliff, depending on light exposure and on the depth and water-holding capacity of the soil.

My field site is 1,700 kilometres from my laboratory. It's a two-hour flight from my lab to the closest airport, in Johannesburg. Then I have to drive for 4.5 hours. In daylight, the rock surface at the site is at least 40 °C, and tools will burn your hand – and I can't always get the equipment and supplies I need.

Last November, I had to drive about 40 kilometres to get dry ice to freeze some samples. I'd forgotten my wallet, but the shopkeeper said I could pay next time. That was incredibly heart-warming.

Rose A. Marks is a postdoctoral researcher in molecular and cell biology at the University of Cape Town in South Africa. **Interview by Abdullahi Tsanni.**

Photographed for *Nature* by
Jennie Lyn Pretorius.